Generalized Volterra-Wiener and Surrogate Data Methods for Complex Time Series Analysis

by

Akhil Shashidhar

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Masters of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2006
[September 2006]

Author .....
Department of Electrical Engineering and Computer Science
August 11, 2006

Certified by .....
Chi-Sang Poon
Principal Research Scientist
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Generalized Volterra-Wiener and Surrogate Data Methods for Complex Time Series Analysis

by

Akhil Shashidhar

Submitted to the Department of Electrical Engineering and Computer Science
on August 11, 2006, in partial fulfillment of the
requirements for the degree of
Masters of Engineering in Electrical Engineering and Computer Science

Abstract

This thesis describes the current state-of-the-art in nonlinear time series analysis, bringing together approaches from a broad range of disciplines including the nonlinear dynamical systems, nonlinear modeling theory, time-series hypothesis testing, information theory, and self-similarity. We stress mathematical and qualitative relationships between key algorithms in the respective disciplines in addition to describing new robust approaches to solving classically intractable problems.

Part I presents a comprehensive review of various classical approaches to time series analysis from both deterministic and stochastic points of view. We focus on using these classical methods for quantification of complexity in addition to proposing a unified approach to complexity quantification encapsulating several previous approaches.

Part II presents robust modern tools for time series analysis including surrogate data and Volterra-Wiener modeling. We describe new algorithms converging the two approaches that provide both a sensitive test for nonlinear dynamics and a noise-robust metric for chaos intensity.

Thesis Supervisor: Chi-Sang Poon
Title: Principal Research Scientist

Acknowledgments

I thank my research supervisor, Dr. Chi-Sang Poon, for his constant support, advice, and efforts toward this work. I also thank the members of my lab and my friends for making my education and research memorable. Finally, I thank my mother Sarala, father Shashi, and sister Deepthi for more than my infinite page limit allows.

Contents

I Nonlinearity and Complexity in Time Series **15**

1 Historical background **16**

2 Deterministic approaches **19**

2.1 Simple invariant sets 20

2.2 Chaos 21

2.3 Embedding and Reconstruction 22

2.3.1 Embedding dimension 23

2.3.2 Embedding lag 25

2.3.3 Alternate embedding methods 26

2.4 Dimension measures of state-space attractors 27

2.4.1 Capacity dimension 27

2.4.2 Correlation dimension 28

2.4.3 Generalized dimensions 33

2.4.4 The attractor dimension as a measure of complexity 34

2.5 Measuring sensitivity to initial conditions using the Lyapunov exponent 36

2.5.1 Calculation from time series 37

2.5.2 Lyapunov spectra 38

2.5.3 Kaplan-Yorke conjecture 39

2.5.4 Meaning of invariant measures 40

2.5.5 The Lyapunov exponent as a measure of complexity 41

3 Stochastic approaches **43**

3.1 Information theory and entropic measures 43

3.1.1 Kolmogorov-Sinai entropy 44

3.1.2 Calculations 45

3.1.3 Pesin's identity 46

3.1.4 Entropic quantification of complexity 47

3.2 Self-similar stochastic processes 48

3.2.1 Monofractal processes 50

3.2.2 Multifractal processes 51

3.2.3 Infinitely Divisible Cascades 53

3.2.4 Multifractal quantification of complexity 53

4 Revised models of complexity **55**

4.1 Stochastic complexity in the NARFIMA model 57

4.2 Deterministic complexity in the NARFIMA model 59

4.3 Complexity as predictability 60

4.4 Revised complexity 61

II Generalized Volterra-Wiener and Surrogate Data Methods 63

5 Surrogate and VAR approaches 64

6 Uses and misuses of surrogate data 66

7 Generalized Volterra-Wiener theory 71

8 Detection and quantification of chaos using noise titration 77

9 Conclusion 81

III Appendices 83

A Nonlinear additive noise reduction 84

B The method of surrogate data for hypothesis testing 88

B.1 Null hypothesis 89

B.2 Discriminating statistic 92

B.3 Statistical testing 93

B.4 Misconceptions 94

B.4.1 Misinterpreting hypothesis non-rejection 94

B.4.2 Misinterpreting hypothesis rejection 95

B.5 A test for nL processes 95

C Nonlinear system modeling using Volterra-Wiener functional expan-

sions **99**

C.1 Volterra Series Representation 100

C.2 Wiener Series Representation 101

C.3 Kernel Estimation 104

D Detection of nonlinear dynamics in short, noisy time series **106**

D.1 Orthogonal-search algorithm for efficient autoregressive system estimation 109

D.2 Implementation of orthogonal search algorithm 111

D.2.1 Estimating α_{mr} 111

D.2.2 Estimating g_m 114

D.2.3 Deriving the Volterra kernels 115

E Sensitive detection of chaos using noise titration **117**

E.1 Choice of noise 118

E.2 Nonlinear dynamics detection algorithm 118

F Results from simulations **121**

F.1 Data generation 122

F.2 Results 126

G External Software **131**

List of Figures

2-1 An embedding for the Lorenz attractor (top) along with the original state-space attractor (bottom). Although the two attractors appear different, they are in fact diffeomorphic. 24

2-2 An example of correlation dimension calculation applied to the logistic map $x_{n+1} = 4x_n(1 - x_n)$. The x-axis is the logarithm of the ball size ϵ and the y-axis is the logarithm of the correlation integral $C(\epsilon)$. The slope of the best-fit line in this case is 0.73, approximately the dimension of attractor of the logistic map. 30

4-1 The multifractal spectra of two machine-generated self-similar stochastic signals are shown. The red signal is governed by $(1 - B)^{0.2}x_t = \epsilon_t$. The second signal is governed by $(1 - B)^{0.2}y_t = \epsilon_{t-1}^2\epsilon_t$. The wider curve for the nonlinear signal reflects its greater stochastic complexity. Note that due to finite data sets, the linear signal also appears slightly multifractal. 58

6-1 The null and alternate hypotheses of the surrogate method with the IAAFT algorithm. In the null hypothesis, the input w is a white Gaussian processes, passed through a linear filter L and a nonlinear static observation function n respectively. The alternate hypothesis captures all other classes of signals by replacing the linear filter L with a nonlinear filter N. 69

6-2 (a) A differentiated χ^2 process with a realization of this process of length 10,000 is shown. (b) Correlation dimension estimates of the series (circle) and surrogates (crosses) using the Grassberger-Procaccia algorithm with $\tau_d = 3$ [51]. (c) Approximate entropies of the series (circle) and the surrogates (crosses) using parameters $m = 2$ and $\tau = 1$ [116]. 70

7-1 The VAR method for detection of nonlinear dynamics compares the predictive power of an optimal linear predictor model with a corresponding nonlinear predictor model (left). The VAR-titration method iteratively adds noise to the signal until nonlinearity is no longer detected (right). Nonlinearity is detected if the cost function of the nonlinear predictor (blue trajectory) is significantly lower than the cost function of the linear predictor (red trajectory). 75

8-1 The sensitivities of the surrogate versus the VAR-titration methods applied to both deterministic autoregressive (solid box) and stochastic moving-average (dashed box) processes. The surrogate method can only reject the null hypothesis of Ln stochastic processes. The VAR-titration method, on the other hand, provides a sensitive test for chaotic dynamics. In addition, the method quantifies the strength (maximal Lyapunov exponent) of the underlying dynamics. 80

A-1 An example of the nonlinear noise-reduction scheme applied to the logistic map $x_{n+1} = 4x_n(1 - x_n)$ with 10% measurement noise. The top graph shows the embedding of the time series without noise reduction, and the bottom graph shows the embedding with nonlinear noise reduction. The dynamical behavior of the system is generally unchanged, unlike linear noise reduction schemes which generally corrupt nonlinear dynamics. 87

B-1 (a) The signal of question. (b) Correlation dimensions of the original series (blue circle) and surrogates (red crosses). (c) Approximate entropies of original series (blue circle) and surrogates (red crosses). 96

D-1 An example of the autoregressive Volterra method for the detection of nonlinear dynamics in a time series, here applied to noise corrupted colored Gaussian process (top) and a noise corrupted chaotic map (bottom). The nonlinear predictor (blue) is performs significantly better than the linear predictor (red) in the chaotic map and thus nonlinear dynamics are detected. The null hypothesis of linear dynamics are not rejected for the Gaussian process. 108

E-1 Noise limit results from titration of chaos (blue) with mathematically calculated Lyapunov exponent values (red) for a range of parameter values for the logistic map. 119

F-1 When discretizing flows, the sampling interval often affects the results. Here we see the VAR p-value calculation results for the Mackey-Glass flow with $\tau = 100$. All series are of length 1000. A p-value less than 0.05 implies detection of nonlinear dynamics. As we see here, a sampling interval too small exaggerates local linear dynamics; whereas, a sampling too large destroys all temporal correlations. 129

F-2 LnL noise added to the logistic map with $r = 3.5$ (with wraparound to prevent divergence). With no dynamic noise, the series is periodic. A small amount of dynamics noise drives the system into chaos; whereas too much masks the nonlinearity. We have found that other moving-average sequences produce similar results. 130

List of Tables

E.1 Noise limit comparisons for noise with various distributions and linear filters. The LPF is a exponential moving-average low-pass filter with $\tau = 10$. 119

F.1 The fractions of null hypothesis rejections from the autoregressive Volterra and noise titration methods applied on a randomized collection of non-linear autoregressive series. Periodicity was tested before the noise was added using an iterative pattern matching scheme on the steady-state components of the time series. 129

F.2 The fractions of null hypothesis rejections from the autoregressive Volterra and noise titration methods applied on a randomized collection of non-linear ARMA series. Periodicity was tested on the MA-free data series. Note that many series autoregressively periodic enter the chaotic regime when driven by dynamical noise. 130

Introduction

Linear approaches for understanding time series span a wide range of fields including systems theory [107, 78], differential equations [60], vector spaces [141], stochastic processes [110], and state evolution [154, 156]. These approaches proved vastly powerful in the understanding of nature and the engineering of systems.

For certain classes of problems, the linear hypothesis proves false and traditional methods are modified to accommodate nonlinear behavior. For instance, the studies of differential equations and state evolution yielded the fields of nonlinear dynamics and chaos [142, 59, 48]. Linear systems theory yielded analogous studies in nonlinear systems [95, 92]. Algorithms for state-space estimation such as the Kalman filter were modified to yield the extended Kalman filter and particle filters [156]. In addition, detection algorithms were designed to decide if nonlinear tools are required for further analysis or if the linear theories suffice [51, 155, 143, 7, 120, 149].

Behavior more intricate than that explained by linear or simple nonlinear descriptions are often deemed *complex*. Although the colloquial definitions are clear, the use of the term in time series analysis is often inconsistent and sometimes contradictory. In Part I, we investigate several disciplines for nonlinear time series analysis with a

focus on approaches to defining complexity rigorously. We then propose a unifying description of complexity encapsulating many of the current approaches.

In Part II (pending publication), we describe the current state-of-the-art in robust nonlinear time series analysis in addition to converging classical methodologies with modern techniques to propose better approaches to quantification and classification. Details of the methodologies used in Parts I and II are described in the Appendices.

Part I

Nonlinearity and Complexity in Time Series

Chapter 1

Historical background

Ancient cultures understood that the past could often predict the future. The Greeks, for instance, calculated and predicted the motion of heavenly bodies to breathtaking precision. Thales of Miletus predicted solar eclipses to within one year's precision around 600 B.C. The mathematics developed for this scientific understanding was also used for engineering tasks. In 1972, a complex differential gear system was found in a sunken ancient Grecian ship possibly used for the prediction of orbits [140].

In Europe, this period of discovery was largely slowed down during the dark and middle ages as societal focus moved from scientific to religious explanations for nature. The rebirth of scientific approaches to characterization and prediction in science occurred during the Renaissance, where influential physicists such as Galileo Galilei and Johannes Kepler created deterministic theories to explain fundamental natural phenomena [44]. Unlike the previous theological arguments, Galileo's approach was not prescriptive, but rather descriptive in form. He did not answer *why* falling objects do so, but was able to describe precisely the nature by which they fall.

Working parallel to Galileo, Kepler developed the mathematics for the descriptions of heavenly motion, painstakingly analyzing the data collected by colleague Tycho Brahe [72]. Galileo's ideas of falling bodies with Kepler's descriptions of heavenly motion were united by Isaac Newton's realization that the same set of gravitational equations could be used to describe both phenomena.

To describe his findings, Newton created a body of mathematics known today as *calculus* [103]. The birth of calculus led to a surge of mathematical and physical advancement including the Leonard Euler's contributions to modern-day analysis [36], Brook Taylor's studies of vibrations [89], Jean Le Rond d'Alembert's work with partial differential equations [27], Joseph Fourier's investigations of heat flow [42], and Pierre Simon de Laplace's discovery of fundamental electric potential theory [79].

The tools developed based on calculus were vast and led to a great range of scientific advancement. However, certain problems were still left unsolved. For instance, the motion of two heavenly bodies could be described by a set of differential equations and solved. The motion of three bodies could be described by a similar set of equations but were unsolved. Even more difficult were problems such as the description of the dynamics of gaseous fluids, requiring the modeling of a huge number of particles with complex dynamics.

An alternate approach to describing natural phenomena was borrowed from Girolamo Cardano, a scholar of gambling [20]. Cardano had systematized the notion of probability for solving gambling problems, and the mathematics of the work was described by Christian Huygens in *On Reasoning in Games of Chance* in 1657 [64]. These probabilistic ideas were first applied to the social sciences as a statistical tool;

prominent examples include Francis Galton's studies of anthropology [45], Ysidro Edgeworth's studies of economics [33], and Karl Pearson's studies in philosophy [113].

In 1873, the venerable James Clerk Maxwell of electromagnetism fame suggested the use of probabilistic measures in the natural sciences:

> The smallest portion of matter which we can subject to experiment consists of millions of molecules, none of which ever becomes individually sensible to us. We cannot, therefore, ascertain the actual motion of any of these molecules; so we are obliged to abandon the strict historical method, and to adopt the statistical method of dealing with large groups of molecules [96].

The statistical or *stochastic* framework soon gave birth to a wide array of fields including Mendelian genetics [98], information theory [134], and systems theory [153].

The two approaches of scientific analysis led to two fundamentally different approaches to time series analysis. The stochastic approach treated the time series as a statistical creature, using probabilistic models to quantify and describe the behavior. An alternate approach, using deterministic tools to explain erratic behavior was popularized after the discovery of *chaos*, or irregular deterministic behavior. Here we describe approaches to time-series analysis from both the deterministic and stochastic points of view. We also stress the idea of *complexity* from these approaches including commentary on how the methodologies differ.

Chapter 2

Deterministic approaches

An ordinary differential equation of the form

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$$

is known as a continuous-time dynamical system (flow) with state vector $\mathbf{x}$. The dimension of the state vector is known as both the dimension of the system and the number of degrees of freedom. One distinguishing characteristic of a dynamical system is that the time evolution of such a system is dependent only on the state vector and not on the absolute time. If the mathematical formulation of the system is *nonautonomous* (with dependence on time), a dynamical system representation often can be formed by adding an extra dimension to the state vector.

Furthermore, if the function f is *Lipschitz continuous* (essentially a finite cap on the first time derivative), the existence and uniqueness of the solutions given an initial condition $\mathbf{x}(0)$ are guaranteed. The precise formulation of this statement is known as

the *no intersection theorem* [59, 60].

In discrete time, a dynamical system takes the form

$$\mathbf{x}_{n+1} = F(\mathbf{x}_n)$$

Discrete-time dynamical systems, or maps, are viewed in two lights. First, they can represent a numerical integration scheme (e.g. forward Euler, backward Euler, Runge-Kutta) applied to a continuous-time dynamical system. The second interpretation popular in dynamical systems theory is that a discrete-time dynamical system inherently represents a dynamical system with discrete behavior. This approach is borrowed from linear state-space evolution theories.

2.1 Simple invariant sets

A *fixed point* $\mathbf{x}_k$ in the state space of a map satisfies $\mathbf{x}_k = F(\mathbf{x}_k)$. Such points are important because they represent a dynamical ending for all trajectories which hit the fixed point. The corresponding condition in continuous time is $\frac{d\mathbf{x}}{dt}|_{\mathbf{x}=\mathbf{x}_k} = 0$ [142].

Fixed points are classified by their stability. The trajectories from points near a stable fixed point, or *sink*, evolve toward the sink. On the other hand, *sources* are unstable fixed points. A slight perturbation from a source will lead to paths leading away from the fixed point. Saddle fixed points source in some directions and sink in the others.

Flows in one dimension are qualitatively governed by sinks and sources, perhaps

at the infinities. As the dimension of the system increases, more interesting invariant geometries arise in the state space. Two-dimensional systems may exhibit closed one-dimensional invariant sets known as limit cycles. Limit cycles are the state space representation of periodic motion in the system. As with fixed points, both stable and unstable limit cycles exist. In general, stable invariant sets are known as *attractors* and unstable invariant sets are known as *repellers.*

A special form of "periodic" behavior is quasiperiodicity. A quasiperiodic signal can be decomposed into a countable number of periodic signals but is not periodic itself. This phenomenon occurs when the ratio between the frequencies of two periodic components of the quasiperiodic signal is irrational. In state space quasiperiodic invariant sets reside on a torus.

2.2 Chaos

A *dissipative* dynamical system has a negative generalized divergence, i.e. the volume of a set of initial conditions contracts as time progresses. Here we focus on dissipative systems because they are completely characterized by their finite attractors.

In nonlinear flows with a state space with three dimensions or greater, a new type of attractor emerges. The *strange attractor* exhibits a fractal structure. Fractals are infinitely self-similar structures, namely behavior is similar at all scales. Consequently, their dimensionality is ill-defined with the conventional notions of dimension. Different flavors of calculating the dimension of a fractal are discussed later. *Chaos* is the description given to irregular behavior of a dynamical system. It is formally

characterized by two properties:

1. **Fractal attractor** - the equilibrium behavior of the system is irregularity, a fractal attractor in state space.

2. **Sensitive dependence on initial conditions** - two initial conditions close to one another on a strange attractor separate exponentially fast until they span the entire attractor.

Nonlinear time-series analysis is largely founded on demonstrating these properties of chaos.

Why chaos cannot occur in flows lower than three dimensions is explained by the Poincare-Bendixson Theorem [136]. The theorem states that for a two-dimensional flow where all trajectories within some region Q stay in Q for all time, the trajectories settle to either a limit cycle or a fixed point. The theorem precludes the possibility of chaotic behavior in two-dimensional flows; however, it does not apply to chaotic maps as even one-dimensional maps can exhibit chaotic behavior.

2.3 Embedding and Reconstruction

Generally speaking, we do not have measurements of all the state variables of a dynamical system at all time $\mathbf{x(t)}$ but rather some scalar measurement of the state vector $q(\mathbf{x}(t))$. Fortunately, for dissipative systems where we are primarily interested in the geometry and dimensionality of the attractors, we can perform *state-space reconstruction* of the attractor [144]. The *embedding theorem* states if the scalar mea-

surement function q fully captures the information provided by at least one state variable (invertible), we can create an attractor in an alternate space which is *diffeomorphic*[1] to the original. The theorem is powerful because many of the nonlinear algorithms we demonstrate later (such as dimension estimates) are invariant under diffeomorphic transformation.

To perform Takens' reconstruction on a discrete-time system, choose two positive integer parameters, an embedding dimension ϵ_d and embedding lag τ. Then for each measurement $q_k = q(\mathbf{x}_k)$, create a reconstructed vector $\mathbf{q}_k = [q_k, q_{k-\tau}, q_{k-2\tau}, \cdots, q_{k-(\epsilon_d-1)\tau}]^\top$. According to the theorem, the reconstructed space of vectors $\{\mathbf{q}_k \forall k\}$ is diffeomorphic to the original state space attractor if the embedding dimension ϵ_d is greater than twice the Hausdorff dimension of the original attractor. To perform the reconstruction on a continuous-time system, τ may be any positive real number.

Takens' theorem follows from a Whitney's embedding theorem which shows by a topological argument that most k-dimensional manifolds can be embedded in a vector space of $2k$ dimensions[2].

An example of the embedding theorem is given in Figure 2-1.

2.3.1 Embedding dimension

Choosing an appropriate embedding dimension for experimental time series is not always clear. If the embedding dimension is chosen too low, the fully unwrapped attractor may not reconstruct. If the embedding dimension is chosen too high, the

[1]Two smooth manifolds are diffeomorphic if there exists a bidirectionally smooth bijective mapping between them [100]

[2]By "most" we mean manifolds which are both smooth and second-countable. In addition, the theorem is generally stated with the embedding in $2k + 1$ space, which is easier to prove.

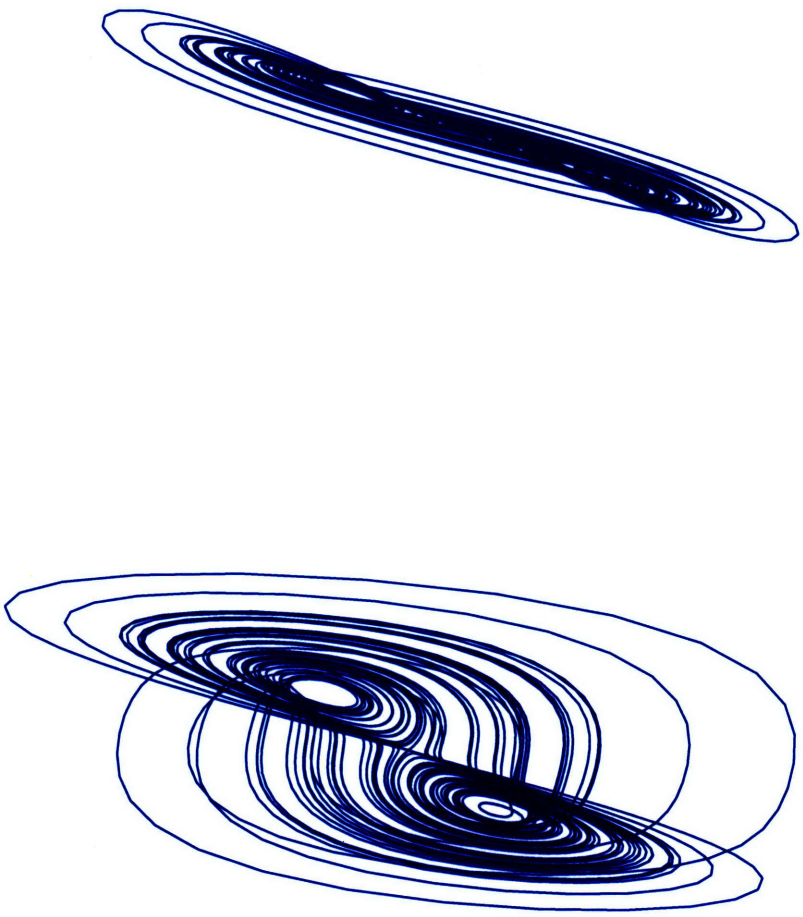

Figure 2-1: An embedding for the Lorenz attractor (top) along with the original state-space attractor (bottom). Although the two attractors appear different, they are in fact diffeomorphic.

finite data may be too sparse in the embedding space to reconstruct properly.

One common estimation technique for the embedding dimension ϵ_d is to use the false nearest neighbors method [71]. This approach estimates an embedding dimension $\epsilon_d = \widetilde{\epsilon_d}$, and for all points $\mathbf{q}_k$ in this embedding, finds the closest few points and calculate the ratios of their distances from $\mathbf{q}_k$. This process is repeated with $\epsilon_d = \widetilde{\epsilon_d} + 1$. The ratios change significantly if the closest points in the original embedding were false neighbors, close only in the projection space. A modified version of this technique has been proposed which claims to be more robust to noisy data [57]

An alternative to finding an optimal embedding dimension directly is to apply nonlinear techniques to a spectrum of embedding dimensions. This method is generally used when the computation time of the respective algorithms is not significant. Generally a plot of statistics over embedding dimensions shows that the statistic converges after a certain embedding. Of course, when the embedding dimension is too high, statistical sampling errors prevail.

2.3.2 Embedding lag

The literature to find the optimal embedding lag τ is much more varied. Although the embedding theorem theoretically applies regardless of the choice of lag, several studies have shown that choosing an appropriate lag value for analysis of experimental signals is important [21, 46]. They show that choosing a lag that is too small will yield spurious artifacts due to autocorrelations. Doing so also amplifies the effects of noise. On the other hand, choosing a lag that is too large may eliminate the dynamic

nature of the signal.

To balance these two issues, many researchers use the first zero of the autocorrelation of the signal as a lag. Fraser and Swinney point out that the autocorrelation is a linear function of the data and is therefore an inappropriate statistic to use for parameterizing a nonlinear technique such as reconstruction. They propose the lag should be the first zero of the time delayed mutual information [43]. A popular alternative to using the mutual information is to use the generalized correlation dimension at different embedding dimensions [83].

2.3.3 Alternate embedding methods

Although the delay reconstruction method is the most popular and prevalent, other embedding techniques exist. A popular choice is the so-called derivative coordinates method: each k-dimensional vector in the embedding space represents the recorded data and its first $k-1$ derivatives. Using finite-difference schemes, constructing the derivative coordinates from a time series is as easy as constructing the delay reconstruction.

This method is often used for reconstruction of flow invariant sets because of the natural mapping between the embedded and real state spaces. One significant problem with the method is that differentiation amplifies the noise component, so for experimental signals, this method is inferior to delay reconstruction. Still, more sophisticated differentiation schemes such as the Savitsky-Golay filter [3] have been used instead of finite differences to alleviate these errors.

[3] A data smoothening algorithm.

Another alternative to delay coordinates reconstruction is the Karhunen-Loeve transformation [14]. The idea is to perform delay reconstruction at a high embedding dimension $\tilde{m}$. Then the covariance matrix of the reconstructed cloud is computed: $C_{ij} = E[\mathbf{x}_i \mathbf{x}_j]$. The eigenvectors of the covariance matrix are then used as the embedding basis.

2.4 Dimension measures of state-space attractors

After reconstructing the attractor, a common follow-up is quantifying the dimension of the attractor. The dimension of the fractal attractor of a chaotic signal is not an integer, so to identify chaos, we may reconstruct the attractor and calculate the dimension of the resulting geometry. In practice, however, tests for chaos using the fractal dimension of the attractor is prone to spurious results due to temporal correlations, short data series, measurement error, and misinterpretation.

2.4.1 Capacity dimension

Regardless, dimension estimates are important from a theoretical and historical perspective and are often used in more modern algorithms. One common dimension measure is the capacity dimension[4]. To calculate the capacity dimension, we choose a scale ϵ and count how many hypercubes of side-length ϵ it requires to completely

[4]The capacity dimension is also known as the cover dimension, grid dimension, or box-counting dimension

cover the fractal. We then take the limit as ϵ becomes infinitesimal:

$$D_{Cap} = -\lim_{\epsilon \to 0} \frac{\log N_\epsilon}{\log \epsilon}$$

where N_ϵ is the number of ϵ-boxes required to cover the fractal. A variation of the capacity dimension is the Hausdorff dimension, which uses boxes of varying sizes to avoid some convergence issues of the capacity dimension. The two terms are often used interchangeably.

The capacity dimension and the Hausdorff dimension suffer from two drawbacks. First, they are extremely time-intensive to calculate. When the embedding dimension is larger than three, these dimension measures are usually intractable [52, 51]. Faster algorithms for calculating the capacity dimension [84] exist, but they are seldom used.

The other drawback to using the capacity dimension is that it calculates the geometry of the attractor, not the invariant set over the attractor. In other words, the capacity dimension does not take into account that some regions of a fractal attractor are more dense than others. Consequently, important information about the dynamics of a system are lost when calculating the capacity dimension.

2.4.2 Correlation dimension

The correlation dimension solves both of these problems. First, the correlation dimension can be calculated directly from the reconstructed vectors in $O(N^2)$ time. Second, the correlation dimension weights dense regions stronger than it weights sparse regions. It is thus more apt for calculating the dimension of a strange attractor. The

two main estimators for correlation dimension from a signal are attributed to Grassberger and Takens. Grassberger's algorithm first calculates the correlation integral for a range of lengths ϵ:

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - ||X(j) - X(i)||)$$

where N is the number of reconstructed vectors analyzed and $\Theta(\cdot)$ is the Heaviside step function. Grassberger uses the Euclidean norm (or 2-norm), but any norm should suffice; latter authors often use the ∞-norm for computational efficiency. The correlation dimension D_{Corr} is then estimated using:

$$D_{Corr} = \lim_{\epsilon \to 0} \frac{\log C(\epsilon)}{\log \epsilon}$$

The correlation dimension of experimental data is usually found by plotting $\log C^m(r)$ against $\log r$ and finding the slope of the best-fit line at the scaling region of the plot. For very large values of ϵ, the correlation integral covers the whole attractor and is thus constant regardless of ϵ. For very small values of ϵ, the correlation integral contains no points other than itself, and is again constant. In between the two extremes, a fractal attractor should exhibit linear or scaling behavior, the slope in this region is the correlation dimension [51]. An example of the correlation dimension calculation can be found in Figure 2-2.

Various optimizations of this technique have been proposed including faster ways to calculate the correlation integral [146] and lessening the number of correlation

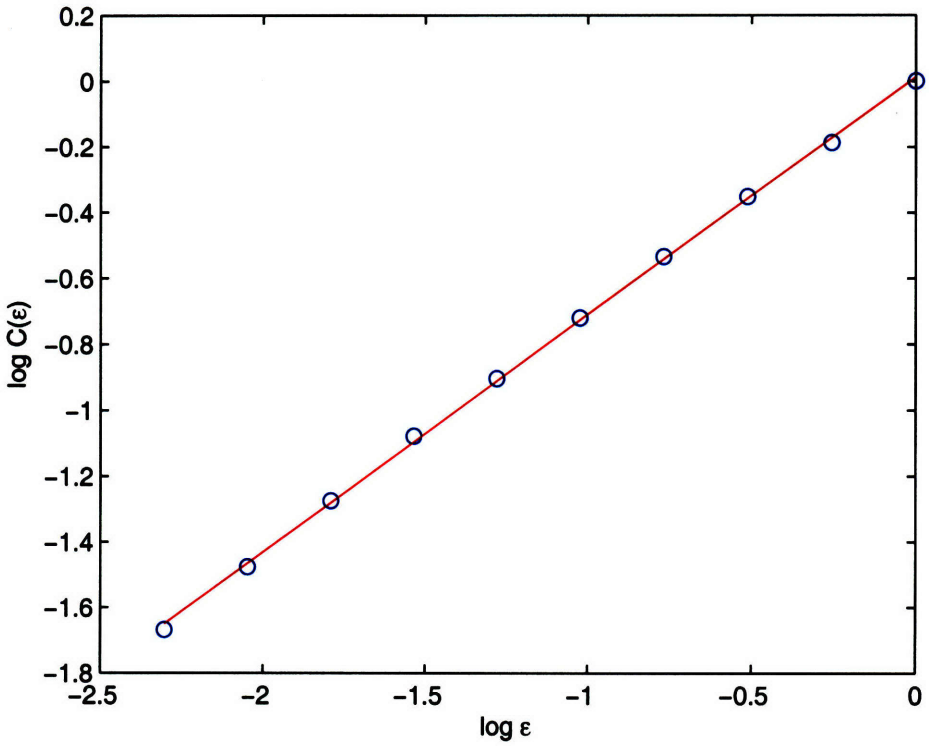

Figure 2-2: An example of correlation dimension calculation applied to the logistic map $x_{n+1} = 4x_n(1 - x_n)$. The x-axis is the logarithm of the ball size ϵ and the y-axis is the logarithm of the correlation integral $C(\epsilon)$. The slope of the best-fit line in this case is 0.73, approximately the dimension of attractor of the logistic map.

integrals calculated [150]. Another useful technique is to use a fast-neighbor search to find the nearby points in the correlation integral. Sophisticated data structures for fast-neighbor searching have been developed in computer science [73, 137]. Schreiber has developed a simplified version of the fast-neighbor search for correlation dimension calculations [130].

Limitations of the correlation dimension

Correlation dimension calculations were used extensively for showing chaotic presence in various time series until it was shown that even some nonchaotic stochastic systems were exhibiting finite correlation dimensions due to temporal correlations in the signal [108]. A commonly used solution to this problem is not counting spatial neighbors which are also temporal neighbors when calculating the correlation integral [148]. Choosing an appropriate cut-off time is solved using a space-time separation plot [123].

Issues other than temporal correlations can complicate the calculation of the correlation dimension. One such issue is rounding or digital precision. Experimental data can be recorded to only limited precision; consequently, all points in the phase space live on some discretized lattice. Solutions to this problem include changing the calculation of the correlation dimension or adding a small amount of noise to the data [99].

Noise, added intentionally or not, proves to be an issue itself when calculating the correlation dimension. Using the Grassberger algorithm, it has been shown that over 2% noise can largely corrupt the results [147]. One means to alleviate this problem

is to change the *kernel function* from $\Theta(\cdot)$ to a smoother function. Diks suggests that using a Gaussian kernel function decreases the method's sensitivity to additive Gaussian noise [30]. An alternative is to apply the algorithm on the reconstructed attractor filtered at different scales [65].

Takens' estimator

Takens' estimator for correlation dimension, a popular alternative to the Grassberger algorithm, takes a probabilistic route by calculating the maximum likelihood estimator for correlation dimension assuming randomly drawn pairs of vectors from the reconstruction have distances distributed normally [145]. It is defined as

$$D_{corr} = \lim_{r \to \infty} \left(-\frac{1}{N_p} \sum_{p=1}^{N_p} \log \frac{r_p(r)}{r} \right)^{-1}$$

where $r_p(r)$ are the distances between randomly chosen points less than r. As we increase N_p, the precision of Takens' estimator improves.

Numerous modifications have also been proposed on this estimator. Borovkova shows the above estimator is biased with respect to the true correlation dimension, and suggests changing the denominator from N_p to $N_p - 1$ will remove the bias and decrease the estimator variance [11]. Ellner showed that Takens' estimators might have problems similar to Grassberger's, namely temporal correlations mistaken for spatial correlations. He proposed an alternative which compensates for this phenomenon by combining the ideas of the Takens' estimator and the chord estimator [34].

2.4.3 Generalized dimensions

We cannot capture the full structure of an attractor (or any fractal) by a scalar. To gain insights to the shape of the attractor, we turn to generalized dimensions, or multifractal spectrum. Just as all the moments of a probability distribution uniquely describe the density function, all generalized dimensions of an attractor uniquely describe the shape of the attractor (up to a diffeomorphism).

Generalized dimensions are defined in terms of the generalized correlation integral:

$$C_q(\epsilon) = \frac{1}{N(N-1)^{q-1}} \sum_{i=1}^{N} \left(\sum_{j=1, j\neq i}^{N} \Theta(\epsilon - ||X(j) - X(i)||) \right)^{q-1}$$

where $p_i(\epsilon)$ is the probability that the i-th hypercube is populated when the space is divided into hypercubes with side length ϵ. Clearly, $C_1(\epsilon) = 1$.

The generalized dimension is then defined as:

$$D_q = \lim_{\epsilon \to 0} \lim_{\gamma \to q} \frac{1}{\gamma - 1} \frac{\log C_\gamma(\epsilon)}{\log \epsilon}$$

It is clear that $D_0 = D_{Cap}$ and $D_2 = D_{Corr}$. The information dimension D_1 is evaluated using the l'Hôpital rule, which requires differentiation of both the numerator and the denominator of the limit. This dimension is so named because the numerator (after differentiation) becomes the average Shannon information needed to specify a point for the resolution ϵ.

Other than the three named dimensions, the other generalized dimensions are usually plotted with respect to q. The curve is non-increasing, and the steeper the

curve, the greater the multifractal behavior. Purely self-similar fractals such as the Cantor set have no q dependence in their generalized dimensions; their multifractal spectra are constant.

Calculating generalized dimensions is not usually a straightforward application of the formula as it was for the correlation dimension. Because of the $q-1$ exponent, small errors in the estimation get magnified, yielding inaccurate results with blind application.

For calculating the generalized dimensions when $q < 2$, the most common method is to used a "fixed-mass" algorithm described by Badii [3]. For $q > 2$, a common method is to reduce the problem to a calculation resembling the conventional correlation dimension [68].

The above description of multifractal attractors should not be confused with the formalism of multifractal signals, a concept popularized by the study of self-similar stochastic processes [Chapter 11]. What is described above is a geometrical description of the invariant attractor set of a nonlinear dynamical system. On the other hand, describing a signal as multifractal implies the geometry of a plot of the signal versus time displays multifractality. The two descriptions are not synonymous.

2.4.4 The attractor dimension as a measure of complexity

One prominent definition of complexity is the attractor dimension in the state space [9, 17, 61, 76, 82, 128, 135]. Using attractor dimension as a measure for complexity is common because such a measure agrees with our intuition about complexity at its

extremes. Intuitively simple and predictable systems such as constant and periodic trajectories have low attractor dimensions. Stochastic systems, on the other hand, have infinite attractor dimensions. We assume from the logic at the end points that the deterministic systems which fall in between the two extrema arrange themselves in a natural "complexity order."

This method suffers from several problems:

- Theoretically, it cannot distinguish between infinite-dimensional attractor spaces such as the partial differential equations of turbulence, the delay-differential equation Mackey-Glass, or a stochastic system such as white noise. Therefore, even in theory, the method can only quantify low-dimensional chaos from higher-dimensional systems and periodic/constant solutions.

- The attractor dimension specifies nothing of the dynamics of the system, or the evolution of points in the basin of attraction to reach the attractor.

- From a practical viewpoint, calculating the correlation dimension estimate of a time-series requires temporally-correlated points to be discarded in the calculation. Any stochastic process with a wide autocorrelation cannot be analyzed using this method without obtaining spurious results. In addition, removing temporally correlated points to obtain meaningful results often requires a huge data set, which is often not feasible.

- Dimension estimates are inherently non-robust. Applying a correlation dimension estimator on anything by the cleanest data set requires careful attention.

2.5 Measuring sensitivity to initial conditions using the Lyapunov exponent

Chaotic dynamics are quantified by both the presence of a strange, or fractal, attractor and sensitive (exponential) dependence on initial conditions. We have described how to quantify the fractal dimension of a strange attractor; here we quantify the other characteristic of chaos - sensitivity to initial conditions.

Exponential sensitivity to initial conditions means two points very close to one another in the state space will separate exponentially fast over time. The rate of this separation is characterized by the *maximal Lyapunov exponent* λ_1. Mathematically, we write:

$$||\mathbf{x}_\tau - \mathbf{y}_\tau|| \sim ||\mathbf{x}_0 - \mathbf{y}_0||\, e^{\lambda_1 \tau}$$

where $\mathbf{x}_0$ and $\mathbf{y}_0$ are two initial conditions close together, and $\mathbf{x}_\tau$ and $\mathbf{y}_\tau$ are their respective time evolutions after τ time units.

A positive λ_1 implies exponential sensitivity to initial conditions. Negative or zero λ_1 implies fixed point or limit cycle dynamics respectively. A stochastic process has an infinite maximum Lyapunov exponent.

Of course, the exponential divergence lasts only until the two trajectories have spanned the attractor. Also, the calculated Lyapunov exponent often varies throughout the attractor. When we speak of the maximal Lyapunov exponent, we refer to an average of the local Lyapunov exponents over the entire attractor.

2.5.1 Calculation from time series

If the underlying dynamical equations are known, standard methods exist to calculate the Lyapunov exponent involving the Jacobian (derivative) matrix. When the dynamical equations are unknown we must use one of several popular algorithms to calculate the maximal Lyapunov exponent from time-series data. The classical method to do so is to start with two initial conditions with a small initial separation and follow their evolutions, rescaling to the initial separation after each time step [155]. The rescaling prevents the entire attractor from being spanned too quickly.

Although this method has been used as a proof for chaos, it suffers from the limitation that stochastic data also register as having positive finite Lyapunov exponents. This effect is because Wolf's algorithm assumes the data demonstrates exponential sensitivity to initial conditions and appropriately fits the model.

A modification of this method compensates for stochastic behavior by checking if the divergence rates are actually exponential [128]. The idea is to choose one point in the state space $\mathbf{x}_{n_0}$ and find all the points in the ϵ-ball around the point $U(\mathbf{x}_{n_0})$. For all the points in $U(\mathbf{x}_{n_0})$, we compute the distance to $\mathbf{x}_{n_0}$ as a function of time separation. This procedure is repeated choosing for all points as $\mathbf{x}_{n_0}$ and averaged:

$$L(\tau) = \frac{1}{N}\sum_{n_0=1}^{N} \ln\left(\frac{1}{|U(\mathbf{x}_{n_0})|}\sum_{\mathbf{x_n}\in\mathbf{U(x_{n_0})}} |\mathbf{x}_{n_0+\tau} - \mathbf{x}_{n+\tau}|\right)$$

The size of the neighborhood ϵ should be chosen to be significantly smaller than the size of the attractor but large enough that each point has a sufficient number of neighbors. The plot of $L(\tau)$ for chaotic systems should have a clear scaling region where

the slope is constant. The slope of this region is the maximal Lyapunov exponent. Stochastic behavior, on the other hand, will demonstrate varying slopes with no clear scaling region.

2.5.2 Lyapunov spectra

A small line segment of points in the state space of a chaotic system will extend in length exponentially fast, the speed characterized by the maximal Lyapunov exponent. If we extend this concept to multidimensional surfaces, we can obtain a whole spectra of Lyapunov exponents. For instance, take a (2-D) disk of initial conditions. The evolution of the disk will be exponentially increasing in one direction; other direction may show either exponential growth or contraction.

For an N-D hypersphere of initial conditions, some directions (eigenvectors) will grow exponentially, while others will shrink exponentially. The corresponding exponents are collectively known as the Lyapunov spectra and are usually ordered from largest to smallest, $\lambda_1, \lambda_2, ..., \lambda_N$. For dissipative systems, the sum of the Lyapunov exponents has to be nonpositive.

Calculation of the entire Lyapunov spectra was also suggested by Wolf by replacing two initial conditions with more. For instance, by creating a triangle of three initial conditions, he shows how to calculate the two largest exponents. This method, however, suffers from the same drawbacks as the corresponding one-dimensional case. It assumes exponential divergence and thus characterizes stochastic signals as chaotic. Because of this problem and others, the method is seldom used.

Rather, a method proposed by Eckmann approximates the Jacobian (derivative) matrix of the dynamics at each point in the trajectory [32]. This is done by finding a linear fit at each point in the state space such that the predicted evolution and actual evolution difference is minimized. With the approximation for the Jacobian, the calculation for the Lyapunov spectra is straightforward through Gram-Schmidt orthogonalization. A modification to this method using polynomial approximations instead of (linear) Jacobian matrices has been proposed [16, 15].

Methods to calculate the entire Lyapunov spectra are generally not as reliable as those to calculate the maximal exponent. This is not usually too large of an issue because the long-term dynamics of the system are characterized by the maximal exponent. Regardless, when calculating Lyapunov spectra, we must be careful to avoiding "spurious" exponents. The cardinality of the spectrum should equal the true dimension of the state space; however, delay reconstruction usually doubles this number. This leads to false, or spurious, exponents in the embedding space.

One method for distinguishing true from spurious exponents is to find the spectra of both the signal and its time reversal [112]. The true Lyapunov exponents of the reversed signal should be the opposite of the true exponents of the original signal. The true exponents can be then found by a matching algorithm.

2.5.3 Kaplan-Yorke conjecture

The spectrum of the positive Lyapunov exponents and the information dimension of the attractor of a dynamical system are hypothesized to be related according to the

Kaplan-Yorke conjecture:

$$D_1 = k + \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|}$$

where k is the number of positive exponents and we assume the exponents are ordered from largest to smallest. This statement has been proven for two-dimensional maps [80]. Even though the exact relation is unproved and several authors have cited high-dimensional counterexamples, the Kaplan-Yorke conjecture provides a good approximation to the information dimension.

2.5.4 Meaning of invariant measures

We may ask why such care is taken to design nonlinear methods such as dimension and Lyapunov exponent measures, when many other nonlinear statistics could be used instead. The reason is that both of these quantities are invariant under linear (and certain nonlinear) transformations. They are hence known as *invariant measures.*

This concept is critical because many of the techniques we apply to time series: embedding, time-frequency analysis, surrogate data, etc. involve rearrangements of the data, but we still need a rigorous quantification of the nonlinear behavior of the data.

Regardless of scale, choice of embedding parameters, or a linear transformation, the Lyapunov exponent and the correlation dimension will remain constant. This is vastly different from most linear metrics of the data, which require the same operations are performed on a collection of signals to accurately compare them.

2.5.5 The Lyapunov exponent as a measure of complexity

Using the maximal Lyapunov exponent as a measure of complexity is attractive for the same reasons as the attractor dimension. Consequently, an estimate for the maximal Lyapunov exponent of a time series has been used as a measure of complexity in the signal [109, 10]. The extremes of the spectrum match the same intuition as above: constant attractors have negative exponents, periodic attractors have zero exponents, chaotic attractors have positive finite exponents, and stochastic processes have infinite exponents.

Unfortunately, the issues with using the Lyapunov exponent as a surrogate for complexity are numerous:

- The type of estimator used for a Lyapunov exponent calculation can drastically affect the result of the outcome, especially in the case of stochastic systems. For instance, Wolf's estimator would assign a relatively low exponent to a brown noise process since the average deviation increases as a square root function of time, the divergence are not even exponential. A more sophisticated estimator, on the other hand would see that brown noise exhibits stochastic behavior and assign a high Lyapunov exponent.

- Calculating the exponent precisely requires reconstruction in the embedding space, which requires a huge number of data points if the embedding is large.

- As with dimension estimates, Lyapunov estimates are not robust. Since estimators are looking for subtle exponential divergences on very small scales, the slightest amount of noise can drastically corrupt an estimate.

Although the extrema of Lyapunov and dimension estimates coincide, the ordering of chaotic series within are certainly not the same. A system with a low-dimensional attractor may easily have a higher Lyapunov exponent than a system with a higher-dimensional attractor. The best we currently have to relate the two approaches is the Kaplan-Yorke conjecture.

Chapter 3

Stochastic approaches

Treating a time series as a realization of a stochastic process rather than a dynamical system brings new tools for analysis. Here we examine two approaches common in the literature, entropic measures and multifractal analysis. Both quantify the time series as a stochastic process, but treat it in different ways. Entropic measures are closely connected to the Lyapunov exponent, as the measure rate of information creation in the stochastic process. Multifractal approaches, on the other hand, draw connections between the geometric structure of the signal and the underlying stochastic processes.

3.1 Information theory and entropic measures

The entropic approach to characterizing time series signals is borrowed from information theory. Generally speaking, an entropy measure quantifies the degree of disorder in the signal. To find the Shannon entropy of a time series, we divide some structure

of the data into boxes of size ϵ.

$$H_{Shannon} = \sum_i p_i \log \frac{1}{p_i}$$

where p_i is the "normalized" weight of the ith box and quantifies a probability. By normalized, we mean that $\sum_i p_i = 1$. The structure chosen is usually either the reconstructed attractor or the data points themselves. The Shannon entropy describes the disorder of the structure; in the first case it describes the "messiness" of the attractor, and in the second case, the disorder of the signal itself. The Shannon entropy is a special case of more general Renyi entropies:

$$H_{Renyi}(q) = \frac{1}{1-q} \log \sum_i p_i^q$$

Through l'Hôpital's rule, we find that $H_{Shannon} = H_{Renyi}(1)$. In practice, the Shannon entropies are more prevalent because of their additivity: the Shannon entropy of a distribution is the sum of the entropies of its mutually exclusive parts.

3.1.1 Kolmogorov-Sinai entropy

The generalized Renyi entropies are useful for quantifying the structure of a probability distribution, or in our case, the structure of the invariant attractor. To quantify the dynamics of a system using an entropic measure, we use the Kolmogorov-Sinai (K-S) entropy. We again divide the attractor into boxes of size ϵ and we define $p_{i_1,i_2,\ldots,i_m}$ as the joint probability that a trajectory goes from box 1 to box 2 through box m.

The K-S entropy is then defined as

$$H_{KS} = \lim_{m\to\infty} \frac{1}{m} \log \sum_{i_1,i_2,\ldots,i_m} p_{i_1,i_2,\ldots,i_m}$$

As with the Shannon entropy, the K-S entropy has generalized versions of higher moments:

$$H_{KS/Renyi}(q) = \lim_{m\to\infty} \frac{1}{m(1-q)} \log \sum_{i_1,i_2,\ldots,i_m} p^q_{i_1,i_2,\ldots,i_m}$$

$H_{KS/Renyi}(0)$ is called the topological entropy because (similar to the Shannon entropy) it does not quantify the dynamics of the system.

3.1.2 Calculations

Calculating the K-S entropy using its definition runs into the same problems we had when calculating the box-counting dimension of an attractor. The number of points required to sufficiently fill small enough boxes is too large for practical calculations. Even if so many points were found, finding the neighbors in the boxes is a computationally intensive process.

One straightforward solution is to return to the correlation integral. We define the correlation integral as earlier, but parameterizing it by embedding dimension rather than box size ϵ, which is fixed:

$$C(m) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - ||X(j) - X(i)||)$$

Then, for very small box sizes, the K-S entropy can be calculated by calculating the

correlation integral in two different embeddings [32]:

$$H_{KS} = \lim_{\epsilon \to 0} \lim_{m \to \infty} \lim_{N \to \infty} \log \frac{C_i(m)}{C_i(m+1)}$$

A practical approximation of the above is to calculate the logarithm under different box sizes for $m = 2$ and find a scaling region. The slope of the curve is the approximate entropy [116].

For very small values of ϵ it has been shown that the approximate entropy calculation is biased because it counts self-matches [127]. Avoiding the self matches when calculating the correlation integral is the solution.

3.1.3 Pesin's identity

The Lyapunov exponents quantify the divergence of initial conditions and the entropy measures quantify the loss of information about the initial conditions. Pesin's identity provides a quantitative link between entropic and Lyapunov measures [115]:

$$H_{KS} \leq \sum_{i=1}^{k} \lambda_i$$

where k is the number of positive Lyapunov exponents. The identity provides an upper bound to the K-S entropy. A common lower bound on the K-S entropy is $H_{KS}(2)$, which can be calculated from the correlation integral.

3.1.4 Entropic quantification of complexity

Unlike dimension or sensitive dependence on initial conditions, entropy is a concept that was applied to stochastic systems before dynamical systems [40, 6, 87, 88, 151, 116]. Because of its probabilistic roots, entropic calculations make more sense and are more discriminatory among stochastic systems than deterministic ones. However, an entropic definition of complexity needs to be scrutinized as thoroughly as the previous approaches:

- *Shannon Entropy.* The Shannon, or information, entropy of a strictly deterministic system is zero because, in theory, the past uniquely defines the future, thus no information is gained as time evolves. An estimate of the Shannon entropy of a time-series, on the other hand, yields much different behavior since it appears chaotic systems follow some probabilistic route, whereas periodic or constant systems do not. The Shannon Entropy is generally not calculated for deterministic signals because its theoretical inability to distinguish between stochastic and deterministic behavior. In addition, the dynamics of the system are not considered, making the information entropy a poor choice for analyzing the complexity of a deterministic system.

- *Kolmogorov-Sinai (K-S) Entropy.* The K-S entropy is more suited for dynamical systems because it quantifies the dynamics or evolution of the trajectories. In a sense, the K-S entropy "embeds" the data in different dimensional spaces. By Pesin's identity, the K-S entropy is closely related to the positive Lyapunov exponents, so using it suffers from the same problems as using the Lyapunov

exponents.

- *Approximate Entropy.* Unlike the above techniques, the approximate entropy is strictly a calculation; it does not attempt to estimate any theoretical quantity. Approximate entropy, we have found, is unable to reliably distinguish between stochastic and deterministic systems. Because of its local predictability, correlated noise generally has an approximate entropy lower than most chaotic time series.

From the above description, it should be clear that entropic definitions of complexity are completely different from the dimensional or Lyapunov definitions.

3.2 Self-similar stochastic processes

Chaos theory introduced the idea that simple deterministic processes can yield irregular, random-like behavior. This does not imply, however, that all irregular signals are generated from a low-dimensional chaotic system. Other methods of analysis of time-series signals exist. For example, many signals, such as the motion of gas particles, result from very high-dimensional systems. Others, such as the turbulent behavior of fluids, are governed by partial differential equations and consequently occupy an infinite-dimensional state space.

The conventional method of modeling such high-dimensional systems is to assume the signal is a realization of a stochastic process [110]. A stochastic process is a generalization of a random variable, with an added time component. A joint probability distribution uniquely classifies the process.

The "simplest" stochastic process is the white Gaussian process, where individual time samples are mutually independent and distributed with identical Gaussian distributions. In the discrete case, this is known as an IID (independent and identically distributed) process. White processes in general have flat power spectral densities. An integrated white process is known as a brown process or "random walk in one dimension" and is characterized by uncorrelated increments.

The Gaussian brown process is often investigated because it displays self-similarity: observed at different time and space scales (zooming in on a graph), a brown noise process still appears brown with its original properties. The signal is similar in all scales: it is a fractal.

It is important to note that this approach is drastically different from the approach described above. Previously, we showed low-dimensional chaos by reconstructing an attractor and showing that the attractor has a fractional dimension. Here, we are showing the path traced out by the signal itself has a fractional dimension. Since the signal occupies a 2-D chart, the dimension of the signal itself lies between 1 and 2. Rather than motivated by dynamical means, we are now motivated by statistical means. Both analyze the structure of the signal in two different senses, but the interpretations are vastly different.

The simple white and brown Gaussian processes have been used to model noise and other simple stochastic systems, but to characterize a wider range of processes, certain generalizations of the two processes are often made. The investigation of such generalizations are collectively known as the theory of self-similar stochastic processes.

3.2.1 Monofractal processes

A Gaussian brown process is known by many names: Gaussian random walk, Gaussian uncorrelated increments process, Gaussian brownian motion, and Wiener process. Whatever name, it is characterized by a covariance function general to all uncorrelated increments processes $K_{xx}(t,s) \propto \min(t,s)$. The standard deviation at a given time is thus $\sigma_x(t) \propto t^{1/2}$.

A generalization of the brownian motion is known as *fractional brownian motion* and is characterized by correlated increments [90]. The standard deviation at a given time is $\sigma_x(t) \propto t^H$. H is known as the scaling exponent or Hurst exponent. If it is greater than $1/2$, the series is said to be persistent – positive correlations between successive intervals. $H < 1/2$ is anti-persistent – negative correlations between successive intervals. The series of intervals of a fractional brownian motion is known as fractional Gaussian noise.

The self-similarity of fractional Brownian motion can be described mathematically: the process $X_\omega(t)$ has the same joint probability density function as a rescaled version $c^H X_\omega(t/c)$ for all rescaling factors c. H is again the Hurst exponent.

The concept of a monofractal process is appealing because a large class of signals can be quantified by one number: the Hurst exponent. Many methods exist to estimate the Hurst exponent of a signal, with the assumption that it is a fractal process. Using such algorithms on non-fractal processes generally yield meaningless results.

A popular algorithm for Hurst exponent estimation is detrended fluctuation analy-

sis [114]. The (brown) signal is divided into segments of some length n. Each segment is *detrended*, its OLS linear component is subtracted. The root-mean-squared value of the detrended signal is computed $F(n)$. This function scales as a power law of the Hurst exponent: $F(n) \propto n^{H+1}$

Detrended fluctuation analysis largely replaced some older techniques of estimating the Hurst exponent such as power spectrum estimation [90], rescaled range analysis [85, 18], and dispersional analysis [8].

Modern methods are generally similar to detrended fluctuation analysis. The scaled windowed variance method, for instance, extends the original algorithm by calculating the standard deviation instead of the root-mean-squared value and suggesting some alternate methods of detrending [19]. An approach that combines several of these measures was proposed recently [126]. A review of some of these approaches can by found in Reference [29].

The monofractal process can also be understood in the context of generalized dimensions with embedding dimension 1. A monofractal process has a dimension independent of q or the moment. Taking the Legendre transformation of C_q will yield a single dot, an alternate approach to identifying monofractal processes.

3.2.2 Multifractal processes

Although the framework of monofractal processes is elegantly parameterized by one number, there exists far more variation within signals than their Hurst exponents. Not all signals are perfectly modeled by fractional brownian motion, and thus monofractal

analysis is always a crude approximation.

As we did above with fractal attractors, we can quantify the dimension of a fractal in a more generic way using generalized dimensions. Instead of assigning one dimension or one scaling exponent to the entire signal, we assign a spectrum to capture more information about the fractality of the signal.

The structure-function approach for calculating the multifractal spectrum was first devised in the context of turbulent fluids [111]. Although popular, the structure-function method suffered from several limitations, and an improved method using the wavelet transform emerged [101]. Muzy et al. showed that the wavelet transform of a multifractal signal $T_x(a,b)$ scaled as a nonlinear function of the order:

$$E[T_x(a,b)^q] \sim e^{aH(q)}$$

The wavelet transform of a purely monofractal process scales linearly with respect to the order, i.e. $H(q) = qH$. Muzy's algorithm uses the maximal lines of the wavelet transform to estimate $H(q)$.

$H(q)$ is a strictly increasing function, and to discern multifractality, we are interested in how far the curve deviates from a line. One popular approach to quantify the deviation is to take the Legendre Transformation of $H(q)$. The transformed signal, the multifractal spectrum, will be an arc, and will be wider the more $H(q)$ deviates from the mean. The wider the spectrum, the more nonlinear the multifractal signal is. Monofractal signals transform to a single point as they are linear.

Unlike the monofractal process, the multifractal process has a generalized dimen-

sion not independent of the moment q. The Legendre transformation of the C_q curve is thus parabolic.

3.2.3 Infinitely Divisible Cascades

Multifractality is governed by the moments of the process behaving as power laws of the scales. If we relax this assumption, a more general class of processes, infinitely divisible cascades, arise [22]. Using the wavelet formalism, an infinitely divisible cascade scales as [2]:

$$E[T_x(a,b)^q] \sim e^{\kappa(a)H(q)}$$

where $\kappa(\cdot)$ is a nonlinear function. For linear choices of $\kappa(\cdot)$, an infinitely divisible cascade reduces down to a multifractal process.

3.2.4 Multifractal quantification of complexity

Using the framework of self-similar stochastic processes, we find a whole new meaning to the term complexity [1, 138, 4, 91]. In this approach, a common interpretation is the width of the multifractal spectrum of a signal quantifies its complexity. This methodology provides a vastly different scale for complexity than the previously described scales.

To begin with, the simplest process are the fractional white and brown Gaussian processes. These are parameterized by a single Hurst exponent H for all orders q and in theory appear as points on the multifractal spectrum. Note that a white process in the other scales of complexity would have appeared at the other end of the spectrum.

As we will show below, wider arcs in a multifractal spectrum suggest more determinism in the signal. In a very broad sense, this definition of complexity is the opposite of the previous definitions.

Using multifractality to quantify complexity applies only to signals that are indeed fractal; they should exhibit some degree of scale invariance. It has been shown that a large variety of experimental signals exhibit some degree of scale invariance, thus multifractal analysis is a quite popular tool for complexity analysis.

Chapter 4

Revised models of complexity

Note that none of these definitions of complexity is necessarily correct. What is universal is that complexity is the opposite of simplicity, and the simplicity of a signal is a function of the axioms used in the analysis. For example, if we approach a white Gaussian signal assuming it was the results of some nonlinear dynamical system, we marvel at its infinite dimension, how quickly nearby points separate, and high rate of new information flow. We naturally conclude the signal is very complex. The same signal approached from a stochastic point of view is not complex at all: jointly Gaussian density function and an impulse autocorrelation.

Here, we create a generalized description of causal complexity which encapsulates the above-mentioned measures. We generalize the idea of an autoregressive moving average (ARMA) model to do so. An ARMA(p,q) model for a time series is given by:

$$x_t = \sum_{i=1}^{p} \psi_i x_{t-i} + \sum_{j=0}^{q} \phi_j \epsilon_{t-j}$$

where $\epsilon(t)$ is a Gaussian white process. This formula may be simplified using the delay operator notation where we use polynomials of lag operator B such that $B^k(x_t) = x_{t-k}$:

$$x_t = \Psi(B) \cdot x_{t-1} + \Phi(B) \cdot \epsilon_t$$

where $\Psi(\cdot)$ and $\Phi(\cdot)$ correspond to the autoregressive and moving average lag polynomials of order p and q respectively.

The above model is strictly linear. We are quantifying complexity of nonlinear deterministic and stochastic signals, so we need to expand our model to fit them. We do so by simply replacing the multiplicative polynomial lag functions with generalized nonlinear versions:

$$x_t = \Psi(B)[x_{t-1}] + \Phi(B)[\epsilon_t]$$

Here $\Psi(B)$ is a nonlinear function of the delays of x_{t-1}, and $\Phi(B)$ is a nonlinear function of the delays of ϵ_t. In addition, we may choose to combine the autoregressive (deterministic) and moving average (stochastic) parts in a nonlinear fashion.

To encapsulate stochastic fractality in the above model, we borrow the fractional derivative operator $(1-B)^\alpha$ from FARIMA models into the moving average component of our model:

$$(1-B)^\alpha x_t = (1-B)^\alpha \Psi(B)[x_{t-1}] + \Phi(B)[\epsilon_t]$$

Given the above model, the free parameters are the nonlinear autoregressive function $\Psi(B)[\cdot]$, the nonlinear moving average function $\Phi(B)[\cdot]$, and the fractional derivative

exponent α. We refer to the above model as a Nonlinear Autoregressive Fractionally Integrated Moving Average (NARFIMA) model.

4.1 Stochastic complexity in the NARFIMA model

If the autoregressive component of the NARFIMA model is negligible, i.e. $\Psi(B)[\cdot] = 0$, the model reduces into long-range self-similar stochastic process with no external inputs. Assume $\Phi(B)[\cdot]$ is linear. Then, the revised model

$$(1 - B)^{\alpha} x_t = \Phi(B) \cdot \epsilon_t$$

and the signal is monofractal. The fractional derivative exponent uniquely determines the Hurst exponent: $\alpha = H + 1/2$ Multifractal scaling arises from a nonlinear moving average function $\Phi(B)[\cdot]$.

As an example, Figure 4-1 shows the multifractal spectra of two signals according to the NARFIMA model with $\Psi(B)[\cdot] = 0$. One signal has a linear $\Phi(B)[\cdot]$ and the other one has a nonlinear $\Phi(B[\cdot])$. The wider multifractal spectra in the nonlinear signal is evident. We choose the fractional derivative exponent $\alpha = 0.2$ for both signals.

As we can see, if the autoregressive component is negligible, the stochastic complexity of a signal is governed by the nonlinearity of the moving average component.

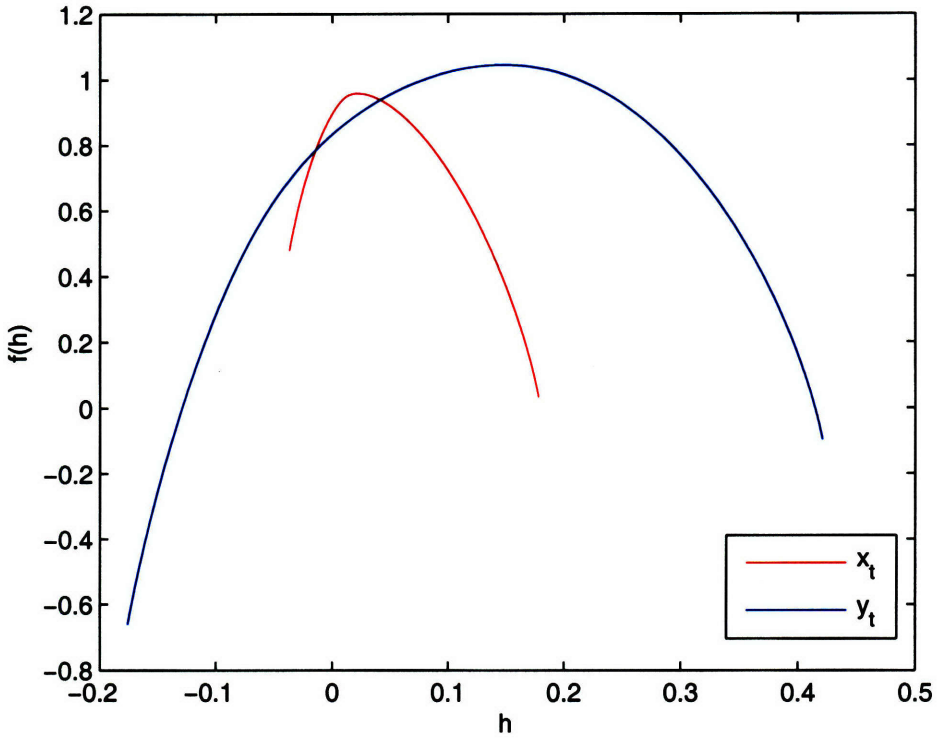

Figure 4-1: The multifractal spectra of two machine-generated self-similar stochastic signals are shown. The red signal is governed by $(1-B)^{0.2}x_t = \epsilon_t$. The second signal is governed by $(1-B)^{0.2}y_t = \epsilon_{t-1}^2\epsilon_t$. The wider curve for the nonlinear signal reflects its greater stochastic complexity. Note that due to finite data sets, the linear signal also appears slightly multifractal.

4.2 Deterministic complexity in the NARFIMA model

When we use the phrase *deterministic complexity*, we refer to the use of the attractor dimension as a complexity measure. The cases of Lyapunov exponents and entropic measures are covered in the next section.

A deterministic dynamical system by definition has no stochastic component. In our NARFIMA model, this corresponds to $\Phi(B)[\cdot] = 0$ such that the model reduces to

$$x_t = \Psi(B)[x_{t-1}]$$

If $\Psi(B)[\cdot]$ is a linear function, the attractors of the system will be either fixed points or limit cycles. A nonlinear autoregressive component is required for chaotic dynamics. The complexity (dimension of the attractor) cannot be immediately discerned from the structure of the autoregressive component. A straightforward approach is the simulate the time series according to the model and calculate a dimension measure (such as correlation dimension) on the resulting output.

Although we do not know the deterministic complexity of the signal from just an autoregressive model, we can say something about the order of the autoregressive polynomial function $\Psi(B)$. The order of the polynomial is akin to the embedding dimension of state-space reconstruction. Thus, we are guaranteed that if we choose a model of order greater than twice that of the attractor in state space, we can reconstruct it without any projective effects.

It may be that the moving average component $\Phi(B)[\cdot]$ is not negligible. In this case, deterministic complexity is difficult to pinpoint. If we truly know there exists a stochastic component, then the attractor dimension is infinite: the trajectory will occupy a structure with dimension equal to the embedding dimension. However, we seldom are given the information that a stochastic component is present in the signal. In these case, results from algorithms that calculate the attractor dimension are generally spurious. As described in the background section, algorithms exist to mitigate the effect of a stochastic component [65] and reduce the effects of temporal correlations [148], but these methods do not truly assign an infinite correlation dimension for stochastic signals.

From a practical point of view, the problem is moot. All experimental time-series have a stochastic component, so they all truly have an infinite attractor dimension. Regardless, applying the correlation dimension algorithm is common nonetheless as a measure of complexity, even if the signal is partly stochastic.

4.3 Complexity as predictability

As mentioned previously, one approach for quantifying the complexity of a strictly deterministic signal is to estimate the Lyapunov exponent, which quantifies the exponential sensitivity to initial conditions. The Lyapunov exponent of stochastic signals is infinite.

Entropic measures are applied to both deterministic and stochastic signals. As with attractor dimension estimates, all experimental time-series contains some stochas-

tic component, so any measure of complexity should account for this. Since Lyapunov and entropic measures are tightly related through Pesin's identity, we may think of them together as one measure of complexity as a means for judging predictability.

We are interested in distinguishing between deterministic and stochastic complexity, but entropic complexity or predictability combines the two together, so it is not quite relevant to our goal.

4.4 Revised complexity

We seek to solve the problem of quantifying both deterministic and stochastic complexity of an experimental time series signal. We propose the following scheme to quantify complexity as defined above.

1. Choose an embedding dimension κ and polynomial order d.

2. Create a nonlinear autoregressive polynomial fit to the data x_t minimizing the Akaike cost function using the chosen parameters [7]. This polynomial is our estimated $\hat{\Psi}(B)[\cdot]$ after the first iteration.

3. Take the observed data and compute the moving average component after the first iteration $\gamma_t = (1-B)^{-\alpha}\hat{\Phi}(B)[\epsilon_t] = x_t - \hat{\Psi}(B)[x_t - 1]$.

4. First check if γ_t is linear Gaussian using the method of surrogate data (with correlation dimension as a discriminating statistic).

5. If γ_t is not linear Gaussian, find the width w_m of its multifractal spectrum to quantify stochastic complexity.

6. Repeat from step 2 with different parameters until $f_c(\kappa, d, w_m)$ is minimized, where $f_c(\cdot)$ is a cost function for complexity.

Note that we do not use nonrobust techniques such as correlation dimension calculation or Lyapunov exponent estimation. In addition, we separate the processes of finding deterministic and stochastic complexity, much unlike measures of entropic complexity.

Part II

Generalized Volterra-Wiener and Surrogate Data Methods

Chapter 5

Surrogate and VAR approaches

The discovery of low-dimensional chaotic dynamics over three decades ago [86, 97] ushered in a wealth of basic analytic techniques that provide systematic characterization of chaotic attractors in deterministic systems [51, 117, 155]. Subsequently, chaotic dynamics have been implicated in many scientific disciplines including psychology [56], physics [77], ecology [26], engineering [24, 5], systems [69], physiology [67], chemistry [35], geology [152], economics [124], sociology [37], and telecommunications [139]. Unfortunately, when the observations are contaminated with noise, many of these classic approaches proved unreliable [108, 128] raising questions about the resultant claims of chaos in experimental signals. To deal with this issue, many techniques have been proposed as possible tests of nonlinear dynamics, or even deterministic chaos, in the presence of measurement noise (e.g. References [143, 147, 128]). Despite this, most of these nonlinear dynamics detection techniques still lack the sensitivity, specificity and robustness necessary for analyzing experimental time series, which are typically short and noisy.

A landmark contribution to experimental time series analysis is the method of surrogate data for linear versus nonlinear hypothesis testing [149, 133]. This method has gained widespread popularity thanks to its simplicity, practicality, and reliability as a general statistical procedure to significantly bolster the efficacy of other nonlinear detection techniques. Along with its celebrated uses, however, come unwitting misuses that pervade the literature reflecting a general confusion among the scientific community at large about the applicability of the method as well as its limits and qualifications.

A fundamentally different approach to nonlinear dynamics detection in experimental time series is the Volterra autoregressive series (VAR) method [7]. In addition to linear vs. nonlinear hypothesis testing, the VAR method provides a sufficient test for chaotic dynamics when used in conjunction with a numerical "noise titration" procedure [120].

Here, we review the pros and cons of the surrogate method and the VAR-titration method and elucidate their theoretical underpinnings from the perspective of the classic Volterra-Wiener theory of nonlinear modeling [95, 92]. The resultant framework uncovers a complementary relationship between the surrogate method and the VAR-titration method. In light of these revelations, we propose a unified approach that combines the best of both these methods to provide a more precise discrimination of linear, nonlinear, and chaotic dynamics than is possible with either method alone.

Chapter 6

Uses and misuses of surrogate data

The surrogate method compares the experimental time series against certain null hypothesis represented by a family of randomly generated "surrogates" of the original series. A common null hypothesis tests if the original series is a realization of some *Ln* process, namely, a linearly filtered Gaussian process measured through some monotonic static nonlinearity [Figure 6-1]. The class of *Ln* processes also includes IID processes and linear Gaussian processes without static nonlinearities. *Ln* processes are completely characterized by their autocorrelation in time domain (or equivalently, the magnitude spectrum in frequency domain) and histogram, which describe the effects of the linear system and the static nonlinearity respectively. For this null hypothesis, a sequential phase randomization-rank ordering algorithm called the Amplitude-Adjusted Fourier Transform (AAFT) was first proposed for generating surrogates with approximately the same autocorrelation and histogram as the original series [149]. To avoid possible distortions in the resultant magnitude spectrum, an improved algorithm (known as IAAFT) [131] iteratively updates the randomized sur-

rogate phase spectrum while keeping its magnitude spectrum and histogram identical to those of the original series.

With an *Ln* process as the null hypothesis, any characteristic measure of nonlinearity or complexity may suffice as a statistic for discriminating the original series from the family of surrogates using standard parametric or non-parametric statistical tests [149, 133]. Common statistics include the correlation dimension or correlation integral [51], nonlinear prediction error [143], maximum Lyapunov exponent [155, 128], approximate or sample entropy [116, 127], time reversibility [31], or higher-order moments [132]. Since different statistics may yield different discrimination powers for any given dataset [132], the ability to reject the null hypothesis depends to some extent on the chosen statistic.

Despite its widespread acclaim, the efficacy of the method is often marred by misinterpretation or overextension of its limits in practice causing numerous false or questionable claims in the literature on a variety of scientific applications. In one extreme, non-rejection of the null hypothesis is sometimes mistaken as proof of linear dynamics [66, 122] or disproof of chaotic dynamics [25, 47]. In fact, failure to distinguish the original series from the surrogates is inconclusive since it could stem from many other reasons such as insufficient or improperly sampled data, excessive noise corruption, or the lack of power in the discrimination statistic chosen [118, 149].

In another extreme, null hypothesis rejection is sometimes misconstrued as evidence of determinism [125] or even deterministic chaos [50, 104]. In fact, surrogates generated by the AAFT or IAAFT algorithms represent only a subclass of linearly correlated processes that conform to an *Ln* structure with Gaussian white input

[Figure 6-1]. This subclass does not include other linearly correlated processes with non-Gaussian input [102]. Non-rejection or rejection of the null hypothesis therefore does not rule in or rule out all linearly correlated processes. As emphasized by Schreiber [131], rejection of the null hypothesis using the IAAFT algorithm "does not imply nonlinear dynamics". To demonstrate the above cautions, we applied the surrogate test (with the IAAFT algorithm) on the time series shown in Figure 6-2(a) generated by differentiating an IID Chi-squared signal. We found that they were indistinguishable from one another with the correlation dimension as discriminating statistic [Figure 6-2(b)]. The seeming indistinguishability of the series from the surrogates is misleading, however, as the null hypothesis was clearly rejected by a variety of other discriminating statistics [Figure 6-2(c)]. This time series is a realization of an nL process which has a reverse structure to the Ln processes of the surrogate hypothesis [Figure 6-2(a)]. Since this process is stochastic, rejecting the null hypothesis is not a test for chaos or even nonlinear dynamics. When applying any algorithmic procedure, one must be mindful of the preconditions of the method. The surrogate method is a rejection test against moving-average Ln processes, but this does not necessarily exclude nonlinear moving-average or other stochastic behaviors.

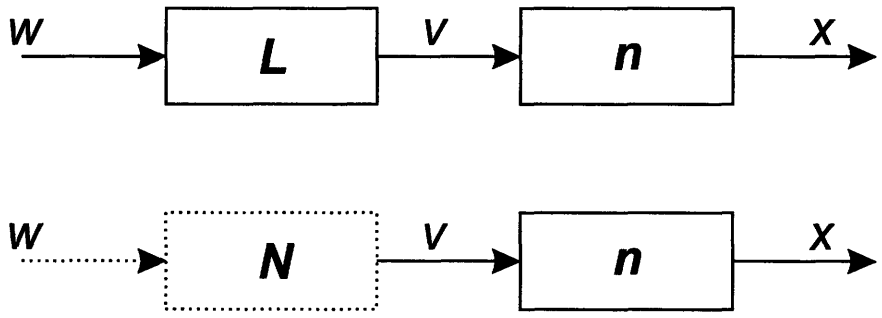

Figure 6-1: The null and alternate hypotheses of the surrogate method with the IAAFT algorithm. In the null hypothesis, the input w is a white Gaussian processes, passed through a linear filter L and a nonlinear static observation function n respectively. The alternate hypothesis captures all other classes of signals by replacing the linear filter L with a nonlinear filter N.

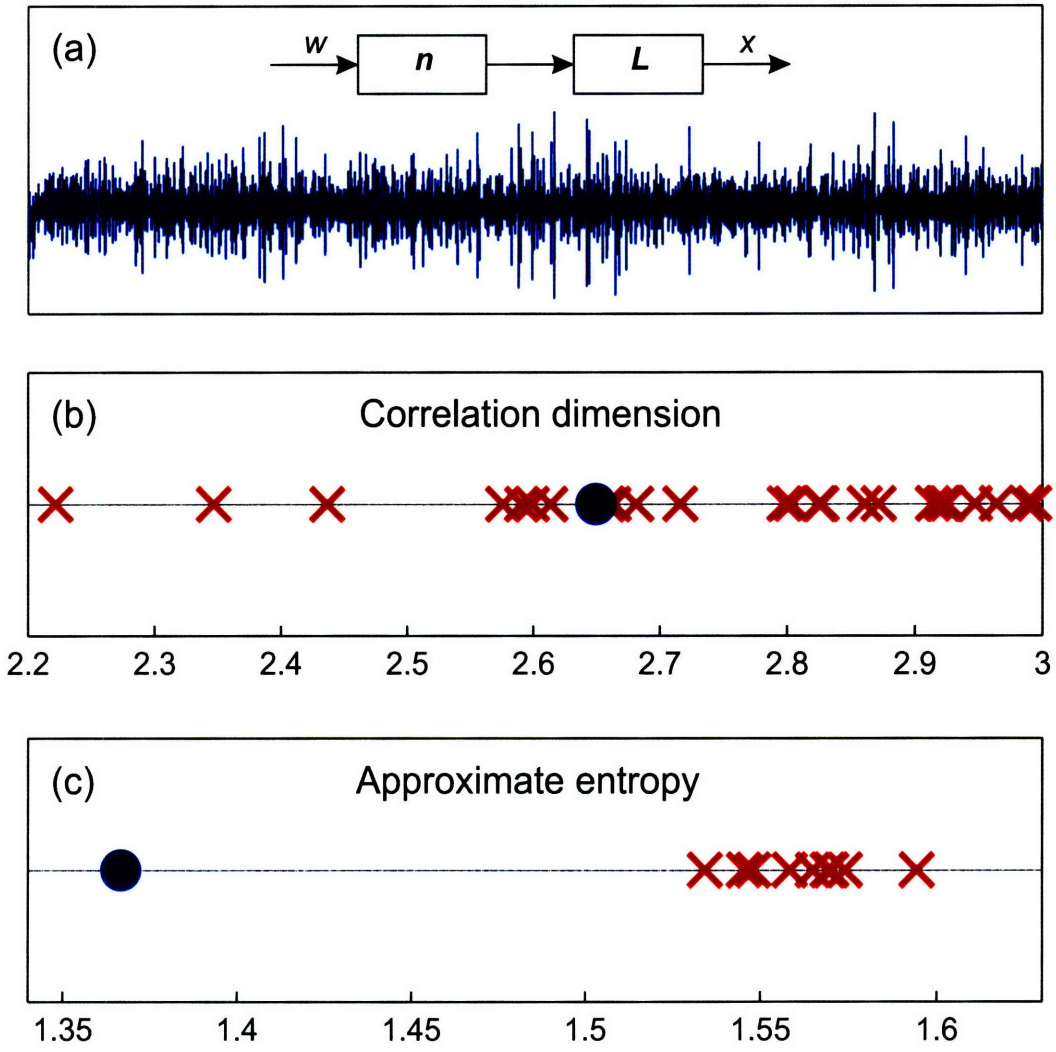

Figure 6-2: (a) A differentiated χ^2 process with a realization of this process of length 10,000 is shown. (b) Correlation dimension estimates of the series (circle) and surrogates (crosses) using the Grassberger-Procaccia algorithm with $\tau_d = 3$ [51]. (c) Approximate entropies of the series (circle) and the surrogates (crosses) using parameters $m = 2$ and $\tau = 1$ [116].

Chapter 7

Generalized Volterra-Wiener theory

Characterization of nonlinear moving-average processes more complex than Ln and nL has been extensively studied in the field of nonlinear dynamic modeling [95, 92], a powerful mathematical theory pioneered by Vito Volterra in the late 19th century and later popularized by Norbert Wiener in the 1930's. The Volterra-Wiener theory has been applied to a wide class of causal nonlinear biological systems including the visual cortex [93], human mechanical dynamics [70], and neurophysiology [129].

The Volterra series expansion formulates the input/output relationship as a polynomial combination of input delays:

$$y(n) = k_0 + \sum_{m_1=0}^{\kappa-1} k_1(m_1)x(n-m_1) + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} k_2(m_1,m_2)x(n-m_1)x(n-m_2) + \ldots$$

where x is the input process, y is the observed process, ki is the set of Volterra kernel

functions, and is a memory parameter. With the assumption of input Gaussianity (as with the surrogate method), the Wiener series representation provides an alternate, but equivalent, set of kernels such that the individual terms are orthogonal, allowing recursive estimation of kernels using a variety of techniques including ordinary least-squares regression [95], covariance calculation [81], fast Cholesky decomposition [74], parallel cascade models [75], coefficient estimation on orthogonal Laguerre basis polynomials [94], iterative cost-minimization methods with neural networks [55, 54], etc.

The Volterra-Wiener modeling approach codifies a large range of nonlinear moving-average processes not captured by the surrogate method. It also provides a means for classifying nonlinear systems in more detail. For instance, the kernels $k_r(m_1, m_2, \ldots, m_r)$ of an Ln system are proportional to $g(m_1)g(m_2)\ldots g(m_r)$ where $g(\cdot)$ is the impulse response function of the linear filter:

$$k_r(m_1, m_2, \ldots, m_r) = \alpha g(m_1)g(m_2)\ldots g(m_r)$$

where α is the rth coefficient in a Taylor expansion of the static nonlinearity n. In contrast, the kernels of an nL system are proportional $g(m)$ on the diagonal and zero elsewhere:

$$k_r(m_1, m_2, \ldots, m_r) = \frac{\alpha}{r}\sum_{j=1}^{r} g(m_j)\delta(m_{\sim j_1})\ldots\delta(m_{\sim j_{r-1}})$$

where $\{\sim j_i\}$ represents the set of all indices not equal to j and α is as above. Such relationships in the kernels are often used to advantage in experimental studies to identify the structure of the nonlinear system in question [63, 70, 129, 95].

Unlike the nonlinear stochastic moving-average processes investigated by the classical Volterra-Wiener theories, chaotic dynamics belong to the class of nonlinear autoregressive processes, where the generating system requires no input signal to create a continuous nonzero output. Traditional methods for detection of nonlinear and chaotic dynamics from time series have proven unreliable in the presence of noise [108, 128]. Sophisticated modern approaches such as the surrogate method also fall short. As shown, the surrogate method in its current form rejects only the *Ln* moving average hypothesis and gives no indication of whether deterministic structure exists in the signal.

In 1996, Barahona and Poon [7] realized that the estimation of system function given just input-output series could be recast to estimate nonlinear dynamics given just a time series. Using the mathematical tools from the Volterra-Wiener theory and statistical hypothesis testing, the Volterra autoregressive (VAR) method provides an easy-to-use and robust test for nonlinear dynamics. The VAR method recasts the moving-average Volterra expansion into an autoregressive form to partially capture the dynamic behavior of the system:

$$\hat{x}(n+1) = k_0 + \sum_{m_1=0}^{\kappa-1} k_1(m_1)x(n-m_1) + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} k_2(m_1,m_2)x(n-m_1)x(n-m_2) + \ldots$$

The method estimates both linear and nonlinear polynomial predictors for a sequence of delay and degree configurations [Figure 7-1]. The maximum delay and degree are chosen for the nonlinear predictor and the maximum delay is chosen for the linear predictor such that both linear and nonlinear predictors have equal number of terms.

The best linear and nonlinear predictors are chosen according to a cost function, balancing a mean-squared error and over fitting penalty for too many terms. If the best nonlinear predictor statistically outperforms the best linear predictor, the null hypothesis of linear dynamics is rejected and nonlinear dynamics are detected. The original paper uses a fast orthogonal search algorithm by Korenberg for kernel estimation [74]. Unlike previous approaches, the Korenberg algorithm yields fast, accurate kernel estimates results on finite and non-Gaussian data with no input or output constraints, making the approach an ideal choice for dynamical system estimation. The optimal nonlinear and linear predictors are chosen using the Akaike cost function, and a Mann-Whitney nonparametric statistical test for comparing the optimal predictors.

The VAR method provides the first published test for nonlinear dynamics in empirical time series. This approach has several advantages relative to previous methods. First, unlike traditional methods, the VAR method does not require the use of surrogate data to ensure its efficacy on noisy time series data. Using the VAR method as a discriminating statistic for the surrogate method abates the power of the method and is thus a misapplication [132]. The method is best applied independently, as better discriminating statistics exist for surrogate testing. A second advantage of the VAR approach is that long data series are not required. From theoretical perspectives, methods such as the correlation dimension estimators converge to the true value as the length of the data approaches infinity. Practically, however, short data series often corrupt and misrepresent true system dynamics. The VAR method requires only enough data to form nonlinear and linear predictors; we have applied the method on time series as short as 500.

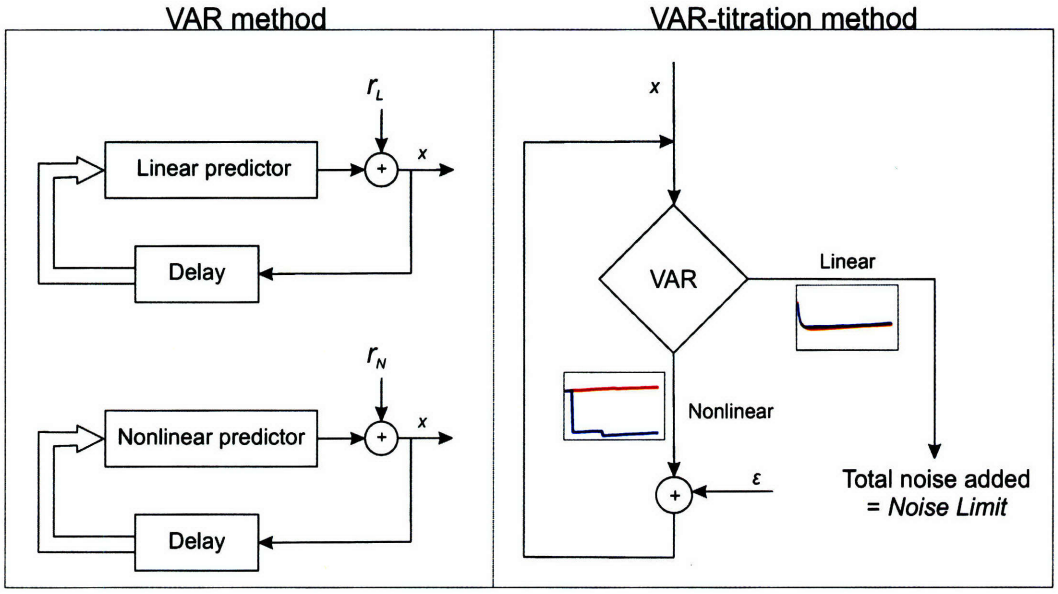

Figure 7-1: The VAR method for detection of nonlinear dynamics compares the predictive power of an optimal linear predictor model with a corresponding nonlinear predictor model (left). The VAR-titration method iteratively adds noise to the signal until nonlinearity is no longer detected (right). Nonlinearity is detected if the cost function of the nonlinear predictor (blue trajectory) is significantly lower than the cost function of the linear predictor (red trajectory).

Finally, the VAR method is robust to measurement noise when the underlying dynamics are chaotic, in stark contrast to traditional estimators such of dimension and Lyapunov exponents. Although classical methods demonstrated nonlinear and chaotic dynamics given zero additive noise, this constraint never exists in experimental data, furthering the need for a sensitive test for nonlinear dynamics and chaos. In the following section, we demonstrate a scheme for detection of underlying chaotic dynamics using the noise robustness of the VAR method.

Chapter 8

Detection and quantification of chaos using noise titration

To test for chaotic dynamics in short and noisy time series, Poon and Barahona proposed a noise addition scheme known as "numerical titration" [Figure 7-1(b)] [120]. If nonlinear dynamics are detected using the VAR method, small amounts of noise are incrementally added until the signal no longer demonstrates nonlinear dynamics. The percentage standard deviation of the original signal at which nonlinear dynamics are no longer detected is known as the noise limit, and quantifies a combination of the sensitivity to initial conditions on the underlying attractor and a measure of the noisiness the system. A positive noise limit signifies detection of chaos. Zero noise limit, on the other hand, implies the underlying nonlinear dynamics follow a periodic or quasiperiodic attractor. With no noise and long data series, recognizing periodic behavior is straightforward, but with short and noisy time series found in experimental conditions, quasiperiodicity or even periodicity are often masked.

In addition to detecting chaos, the noise limit mirrors the maximal Lyapunov exponent of the system dynamics [120]. Some signals are either corrupted by too much measurement noise or the Lyapunov exponent is sufficiently small that the noise limit is very small (e.g. 0.5% standard deviation). Therefore, prior to testing for chaos, it is useful to set a threshold below which we consider to be nonchaotic. We have found that a threshold of 1% for white Gaussian noise provides sufficient sensitivity to very weak chaos (extremely low Lyapunov exponent) while still ensuring no false positives. This threshold applied equally to several types of chaotic maps and flows in addition to different data lengths (ranging from 1000 to 100,000). Noise titration may also use colored or non-Gaussian noise. In such cases, the threshold may need to be modified. In general, the threshold for chaotic detection may be analyzed statistically once a cohesive mathematical framework for the method is constructed. In the current stage, however, the VAR-titration method is understood only heuristically and confirmed empirically. Consequently, the precise threshold can only be chosen after experimental testing.

To compare the surrogate method against the VAR-titration method, we applied both on a wide variety of time series of length 1000, spanning the types shown in Figure 8-1. Fifty surrogates were generated with approximate entropy and time reversibility as discriminating statistics. A z-score greater than 3 rejected the null hypothesis. The VAR method used a maximum delay of 6 and maximum degree of 3 and the signals were titrated with 1

Figure 8-1 shows that VAR-titration method provides a sensitive test for chaotic dynamics. The surrogate method, on the other hand, tests for a much broader range

of both chaotic and nonchaotic processes. Similar results were obtained on signals with 10% measurement noise. In this case, the detection rate of both algorithms decreased slightly, but the majority of the signals were detected correctly. In addition to providing a sensitive test for chaotic dynamics, the VAR-titration method quantifies the intensity (maximal Lyapunov exponent) of the underlying chaotic dynamics. By the nature of noise titration, an increase in measurement noise decreases the noise limit.

The VAR-titration method surpasses previous methods of chaos detection in many respects. First, since the VAR method is used for nonlinearity detection, short time series may be analyzed accurately. Second, the method is applicable to nonstationary processes; the original paper detected chaos in an ecological model with a strong periodic component. More important, however, is the method's direct applicability to very noisy time series. Since experimental time series already contain additive noise, further addition is not necessary for detection of chaos. Such processes are said to be autotitrated. The VAR-titration can only tolerate as much noise as the noise limit itself. Chaos will not be detected, for example, in signals with measurement noise greater than the inherent noise limit of the underlying dynamics. Unlike other methods, however, this phenomenon is quantifiable since the threshold for false negatives is equal to the noise limit of the signal.

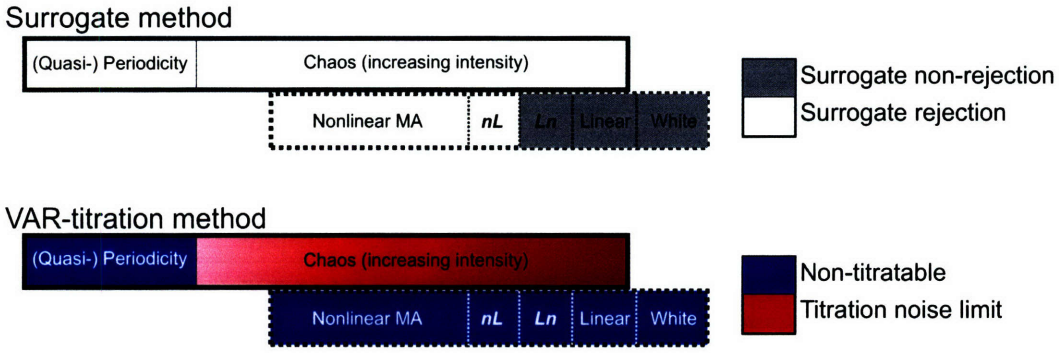

Figure 8-1: The sensitivities of the surrogate versus the VAR-titration methods applied to both deterministic autoregressive (solid box) and stochastic moving-average (dashed box) processes. The surrogate method can only reject the null hypothesis of *Ln* stochastic processes. The VAR-titration method, on the other hand, provides a sensitive test for chaotic dynamics. In addition, the method quantifies the strength (maximal Lyapunov exponent) of the underlying dynamics.

Chapter 9

Conclusion

With the VAR-titration method, we are able to algorithmically identify the presence of chaotic dynamics given only noise-corrupted time series. In addition, the method both identifies nonchaotic nonlinear dynamics and quantifies the initial-value sensitivity of chaotic nonlinear dynamics. If nonlinear dynamics are not detected, the method of surrogate data provides a sensitive test for characterizing the moving-average signal for certain structures. Together, the methods provide detailed characterization of a wide span of nonlinear behavior in time series. Correct and noise-robust characterization of nonlinearity provides a launch pad to both practical and theoretical advances.

The idea of linear-vs-nonlinear hypothesis testing of experimental time series was popularized by the surrogate method and extended by the VAR-titration method. Together they provide a unified framework for the classification of the underlying dynamics of the time series. The surrogate method may be used to search for nonlinearity, in autoregressive or moving-average form. The VAR-titration method may

then characterize the nonlinearity with detection and quantification of nonlinear and chaotic dynamics.

One important direction for further investigation is a thorough analysis of detection algorithms for a class of nonlinear processes with both autoregressive and moving-average (ARMA) components. Such processes may exhibit exponential sensitivity to initial conditions as deterministic autoregressive processes, but also exhibit probabilistic structures due to the moving-average component. Surrogate and nonlinear modeling methods show promise for moving-average (MA) alone, while VAR-titration methods and future improvements help elucidate the autoregressive (AR) aspects. There still, however, exists a void for a suite of practical tools designed for specific detection and analyses of systems with both AR and MA characteristics. Theoretical underpinnings and modifications of classical algorithms on such processes are introduced in a recent text [23]. The VAR-titration method detects the presence of autoregressive dynamics in such processes if the moving-average component is sufficiently small but does not distinguish ARMA from simple AR nonlinear dynamics. Methods for solving such problems may be generalizations of deterministic tools from nonlinear dynamics including surrogate and VAR-titration methods, or may require new theories altogether.

Part III

Appendices

Appendix A

Nonlinear additive noise reduction

Calculating the dimension of the attractor and the sensitivity to initial conditions are important tools in time-series analysis and were the primary means of classifying and quantifying chaos in the late 1980's and early 1990's. They work well when applied to computer-generated chaotic dynamical systems. They also clearly demonstrate chaotic behavior in certain experimental data sets governed by simple nonlinear differential equations under noise-free conditions [41, 62].

When the analyzed signal contains a nontrivial noise component, as most experimental signals do, dimension and Lyapunov measures are unsuitable. Calculating the correlation dimension on a noisy chaotic signal will yield no clear scaling region because the dimension will be too high, thus the embedding will be insufficient. The Lyapunov exponent of noise is infinite, so analyzing noisy data using Lyapunov measures is risky. Generally speaking, these conventional algorithms should be used on data where noise is known to be minimal.

One natural solution to this problem is to use a method of noise reduction for

nonlinear signals. If we use a method that is sufficiently sensitive to noise and the underlying deterministic behavior is strong, we may be able to remove the noise component entirely and analyze the signal using conventional techniques.

Here we focus on additive, or measurement, noise. The other flavor, dynamical noise, introduces the concept of stochastic chaos [23], and is a separate study. Regardless, it has been shown that for a large class of systems, there exists a mapping between additive and dynamical noise [12].

We formulate the problem similar to the one we had earlier. The dynamical system is governed by a state vector $\mathbf{x}$ and an evolution function F:

$$\mathbf{x}_{n+1} = F(\mathbf{x}_n)$$

Instead of measuring a scalar function of the state variable q, we record noise-corrupted version ζ:

$$\zeta_n = q_n + \eta_n = q(\mathbf{x}_n) + \eta_n$$

We assume after embedding, the embedded state vector $\mathbf{q}_k = [q_k, q_{k-\tau}, q_{k-2\tau}, \ldots, q_{k-(\epsilon_d-1)\tau}]^\mathsf{T}$ evolves in embedding space by an evolution function G:

$$\mathbf{q}_{n+1} = G(\mathbf{q}_n)$$

A conventional technique for nonlinear noise reduction (assuming one had a model for

the evolution function $\hat{G}$) is to set all residuals to zero and solve the system [53, 39]:

$$|\hat{\mathbf{q}}_{n+1} - \hat{G}(\hat{\mathbf{q}}_n)| = 0 \ \ \forall n$$

If we have less faith in our approximation of the embedded state evolution function $\hat{G}$, Davies suggested it may be better to simply minimize the mean-squared error [28]:

$$\arg\min \sum_n |\hat{\mathbf{q}}_{n+1} - \hat{G}(\hat{\mathbf{q}}_n)|$$

Davies' paper suggests using the method of gradient descent to minimize the expression. The above method still assumes the existence of an approximation to the evolution function G.

If it is infeasible to approximate an evolution function in the embedding space, a common alternative technique called local projective noise reduction is often used [32]. This method approximates the evolution at each point as a linear prediction.

An example of the above method is given in Figure A-1.

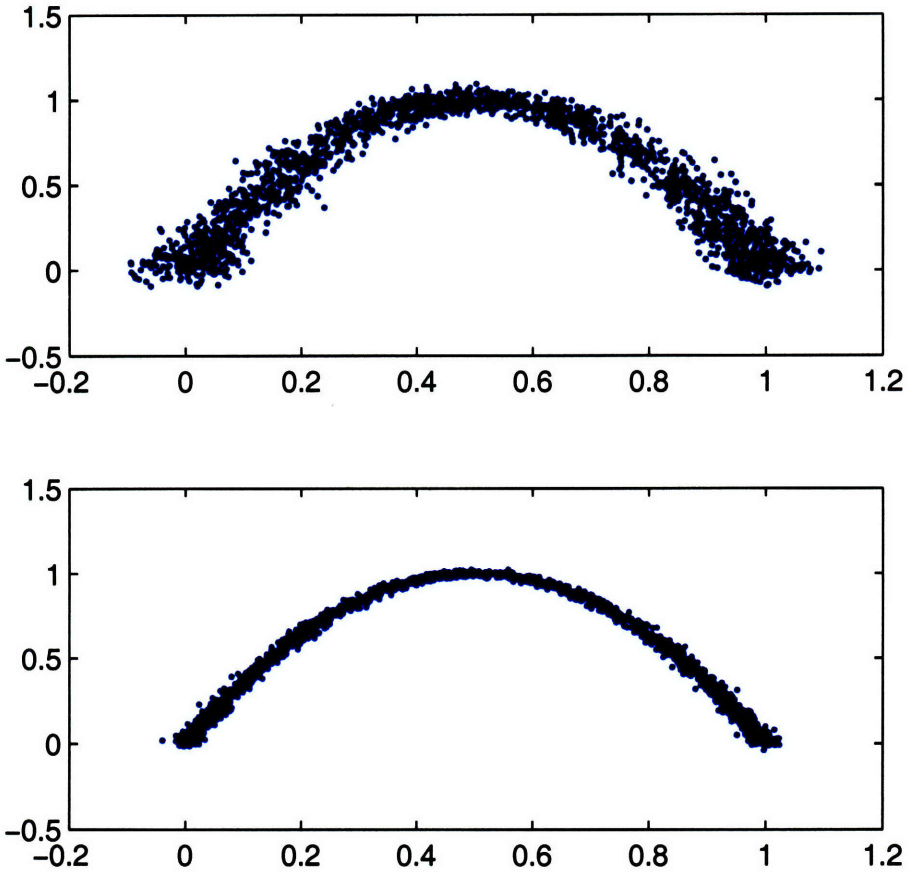

Figure A-1: An example of the nonlinear noise-reduction scheme applied to the logistic map $x_{n+1} = 4x_n(1 - x_n)$ with 10% measurement noise. The top graph shows the embedding of the time series without noise reduction, and the bottom graph shows the embedding with nonlinear noise reduction. The dynamical behavior of the system is generally unchanged, unlike linear noise reduction schemes which generally corrupt nonlinear dynamics.

Appendix B

The method of surrogate data for hypothesis testing

Nonlinear noise reduction is an often useful technique, but in many cases, the noise is too strong or too complex to reduce reliably. A completely alternative approach for time series analysis known as the method of surrogate data and presents a systematic approach to testing hypotheses on a signal in the presence of noise [149].

1. Create a null hypothesis to test in the experimental signal. Assume that the experimental signal satisfies this null hypothesis.

2. Create k surrogate signals such that each is a realization of the same process as the original but in all other senses random.

3. Calculate some scalar *discriminating* statistic on all k surrogates and the original signal.

4. If the discriminating statistic of the original signal is significantly different from the statistics of the surrogates, the null hypothesis from step 1 can be rejected.

B.1 Null hypothesis

In theory, the method of surrogate data can test any given null hypothesis on the data given a method of generating the surrogate data. The null hypotheses used in practice, however, are relatively restrictive. The simplest null hypothesis is that the signal is drawn from an independent and identically distributed (IID) process. This structure may be thought of as equivalent to passing a white Gaussian signal through some static (memoryless) nonlinearity. We represent this static nonlinearity with a lower-cased n.

Any given n process is completely characterized by the probability density function of any sample. IID processes are ergodic in all moments; consequently, we can estimate the density function of the process by plotting the histogram of the signal values.

To satisfy the surrogate criterion for the n hypothesis, we need to ensure the histogram of the surrogate matches the histogram of the original signal, but in all other respects is random. One simple solution is to randomly shuffle the original signal to form the surrogates.

A more general null hypothesis is that the signal is an LTI-filtered (colored) Gaussian process. Since such a process is the output of white Gaussian noise passed through an linear (time invariance is assumed), it is known as the L null hypothesis.

Such a process is both Gaussian and wide-sense stationary and is thus completely characterized by its autocorrelation function or equivalently magnitude spectrum. To generate surrogates for this null hypothesis, we seek to preserve the magnitude spectrum, while randomizing all other aspects of the signal.

The magnitude spectrum and the phase spectrum completely characterizes the time-domain signal. Since the magnitude spectrum encapsulates the effects of the linear system, the experimental outcome of an individual realization is captured in the phase spectrum. To generate surrogates for the L null hypothesis, we convert the original signal into frequency domain, shuffle the phase components, and revert back into time domain. While shuffling, it is important the phase remains odd to ensure a real time-domain signal.

A more general null hypothesis is that the signal linear Gaussian process, measured through some static nonlinearity (assumed to be invertible to prevent information loss). Such an Ln process is fully characterized by a complete description of both the L and the n sections. The L subsystem is characterized by the autocorrelation of its output, while the n subsystem is characterized by the histogram of its output relative to its input.

The first algorithm to test for the Ln hypothesis uses the fact that the output of the L subsystem is known to be Gaussian. To generate the surrogate, it uses this fact to invert the n, phase randomize, and then reapply the n component. This method is known as the Amplitude-adjusted Fourier transform (AAFT) and consists of three steps:

1. **Gaussianization.** Generate a white Gaussian signal w the same length as the original signal x. Reorder w to yield ξ such that ξ has the same *rank ordering* as the x. For example, if the kth largest element of x is in the ith position, the kth largest element of ξ should also be in the ith position. The correlated Gaussian signal ξ serves as an estimate for the output of the L subsystem.

2. **Phase randomization**. In the Fourier domain, shuffle the phase components (again ensuring the phase remains odd). Return to time domain to yield $\hat{\xi}$.

3. **Histogram restoration**. Reorder x such that it has the same rank ordering as $\hat{\xi}$. This step reconstructs the static nonlinearity n.

The AAFT sequence attempts to match both the histogram of the output signal and the power spectrum of the intermediate signal. Schreiber shows that the power spectrum of the surrogate signal is generally flatter than the power spectrum of the original signal, leading to spurious results in constructed cases. He proposes an iterative solution to match both the power spectrum and histogram at the output instead of matching the power spectrum at the intermediate step:

1. **Initialization**. Set ξ_1 as a shuffled version of x, the observed signal.

2. **Power spectrum correction**. Calculate the Fourier transform of ξ_1. Replace the magnitudes of the transform with the Fourier magnitude of x, while keeping the phases intact. The inverse Fourier of the result is ξ_2.

3. **Rank ordering correction**. Reorder ξ_2 to follow the same rank ordering of x to yield ξ_3. For example, if the kth largest element of x is in the ith position,

the kth largest element of ξ_3 should also be in the ith position.

4. **Iteration**. Compare the power spectrum of x with the power spectrum of ξ_3.

   (a) If the difference is non-negligible, let $\xi_1 = \xi_3$ and go to step 2.

   (b) If the difference is negligible, the resulting signal ξ_3 is the surrogate.

This approach matches the power spectrum and the histogram repeatedly until the iteration converges.

B.2 Discriminating statistic

Choosing an appropriate discriminating statistic is the other half of using the method of surrogate data appropriately. Interestingly, as long as the statistic is nonlinear, it will apply to all the surrogate algorithms described above. A successful statistic has to measure some aspect of the signal not different among different realizations of the null hypothesis. For instance, a null hypothesis of a linear process would require a discriminating statistic with some element of nonlinearity. Using a linear statistic such as a first- or second-order moment, for instance, would yield spurious results.

Commonly chosen statistics include the correlation dimension estimates, Lyapunov exponent estimates, and nonlinear prediction errors. Since surrogates are generally nondeterministic, it is unwise to read too much into the dimension or prediction error estimates of them. Rather, they should be used simply as statistics for hypothesis testing.

Another class of discriminatory statistics involves taking higher-order moments of

the data. One popular statistic is described by Diks quantifies the time-reversibility of the signal [31]. A simplified version of this statistic $E[x_n x_{n+1}^2 - x_n^2 x_{n+1}]$ is often used instead. These statistics are generally more computationally efficient.

We have found that entropic statistics such as the Approximate Entropy [116] and Sample Entropy [127] yield very reliable results with the surrogate data method. The results from using entropic measures are generally very clear and reproducible.

B.3 Statistical testing

If the signal is sufficiently noise free and the discriminating statistic is sufficiently powerful, demonstrating nonlinearity is trivial and simplistic statistics may be used. One simple example often used is the Z-score, which calculates the number of standard deviations the original signal is from the surrogates, assuming a Gaussian distribution of the statistic. It is calculated as follows:

$$\zeta = \frac{|Q_x - E[Q_s]|}{\sqrt{Var[Q_s]}}$$

where ζ is the Z-score, Q_x is the statistic of the original signal and $\{Q_s\}$ is the set of the statistics of the surrogate signals. Often, a Z-score greater than 3 rejects the null hypothesis.

B.4 Misconceptions

For reasons including its omnipresence and simplicity, the method of surrogate data is often misunderstood and consequently misrepresented. Here we show two common errors in the literature using the IAAFT algorithm and testing for the *Ln* null hypothesis.

B.4.1 Misinterpreting hypothesis non-rejection

Non-rejection of the hypothesis is often misrepresented. If the null hypothesis is not rejected, no claim about the original signal can be made. If the signal truly does not satisfy the null hypothesis, the method of surrogate data may not be able reject the hypothesis for other reasons including, (a) lack of power in the discriminating statistic, or (b) the signal may be too noisy. Unfortunately, these other reasons are often ignored, leading researchers to assume non-rejection is equivalent to acceptance of the hypothesis [66, 122].

If the hypothesis is not rejected, it is incorrect to form *any* conclusion. For instance, one prominent example of this error asserts non-rejection implies no underlying nonlinearity and thus no chaos [25, 47]. As mentioned, non-rejection happen for several other reasons.

The signal in Figure B-1(a) demonstrates this claim by applying the IAAFT procedure on it. Creating 20 surrogate signals and applying the correlation dimension estimator on all 21 signals.

The original signal is indistinguishable from the surrogates [Figure B-1(b)]. It is

incorrect, however, to assume that the null hypothesis is true. Changing the discriminating statistic to the approximate entropy shows the null hypothesis should in fact be rejected [Figure B-1(c)].

B.4.2 Misinterpreting hypothesis rejection

Null hypothesis rejection is often misinterpreted as determinism [125], or the output of some dynamical system, perhaps perturbed by noise. The signal in Figure B-1 is a realization of a differentiated $\chi^2_{(1)}$ IID process and an example of a nondeterministic (stochastic) signal for which the null hypothesis is rejected.

The null hypothesis is rejected because signal is an nL process, not an Ln process. An nL process is mathematically different from Ln process and are thus not interchangeable [Chapter 12]. A special case of the preceding error is assuming null hypothesis rejection implies chaos [50, 104].

The surrogate method is a powerful and sensitive technique for testing for nonlinearity - rejecting processes not Ln. Misapplications of various forms should be questioned.

B.5 A test for nL processes

The original surrogate data method did not propose a test for nL processes, but one can be considered as an extension of the AAFT method for Ln processes. The AAFT method uses a rank-ordering technique to invert the static nonlinearity. From the inverted signal, the standard phase-randomization surrogate approach may be

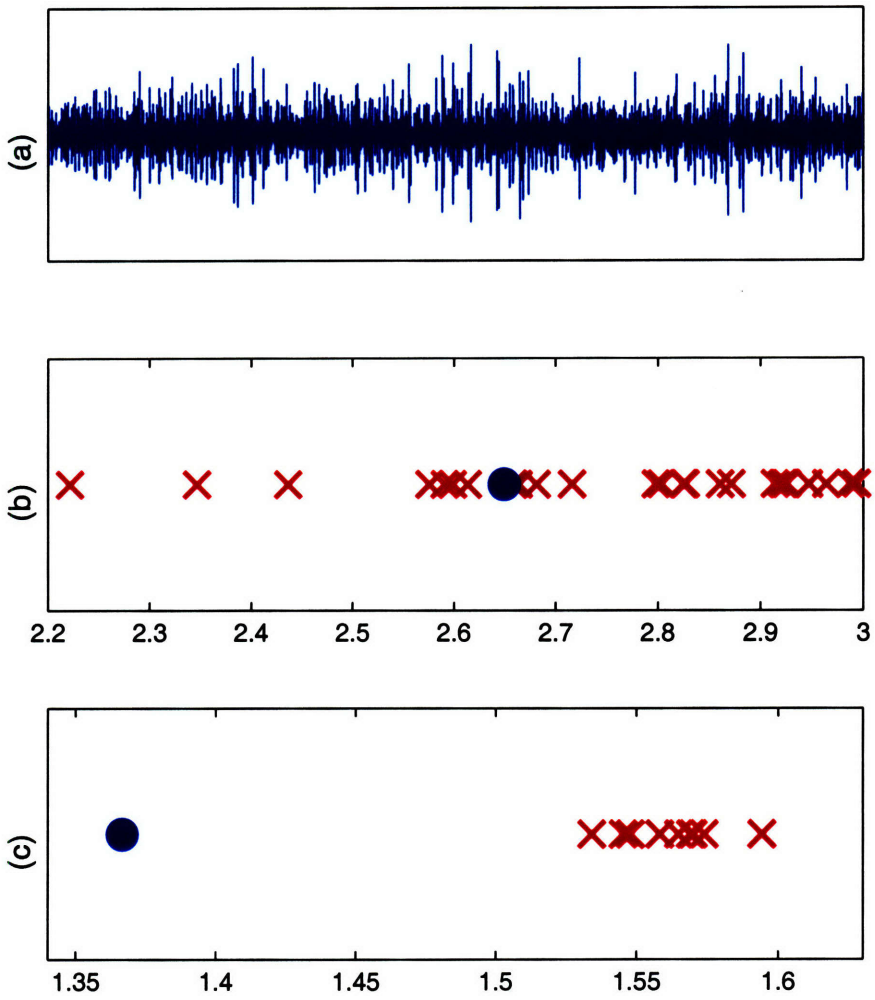

Figure B-1: (a) The signal of question. (b) Correlation dimensions of the original series (blue circle) and surrogates (red crosses). (c) Approximate entropies of original series (blue circle) and surrogates (red crosses).

employed to test for Gaussianity.

Similarly, we propose a surrogate method for the testing of nL processes by first inverting the linearity. From linear systems theory [107, 78], we know that the residuals of a linear predictor with sufficient memory should be approximately independent. Thus the test takes two steps. First, a linear predictor of the data is constructed using ordinary least-squares and the residuals are calculated. The residuals are then tested for independence.

The method of surrogate data provides a robust test for independence using time-domain shuffling to create the surrogate signals. Alternate tests exist, such as the Brock-Dechert-Scheinkman (BDS) test for independence based on the correlation integral [13, 23]. The BDS statistic exploits the fact that the correlation dimensions of independent data for different embedding dimensions are intimately related, namely:

$$\begin{aligned} C(r,m) &= \Pr\{|\epsilon_{tm} - \epsilon'_{tm}| < r\} \\ &= \Pr\{\max_i |\epsilon_{t-i} - \epsilon'_{t-i}| < r, 0 \le i \le m-1\} \quad \text{(B.1)} \\ &= \prod_{i=1}^{m} \Pr\{|\epsilon_{t-1} - \epsilon'_{t-i}| < r\} \quad \text{(B.2)} \\ &= C^m(r,1) \quad \text{(B.3)} \end{aligned}$$

The BDS statistic estimates the difference between the correlation dimensions of some embedding versus an embedding of 1 $C(r,m) - C^m(r,1)$ and tests to see if its statistically nonzero.

Alternate approaches to the linear predictive code were also used. One approach we investigated assumed the linear filter satisfies the minimum-phase requirement,

namely all poles and zeros are within the unit circle. This ensures that both the filter and its inverse are both causal and stable [106]. Such filters exhibit a one-to-one relationship between the magnitude and phase spectra. Since we know the input is IID and thus white, the magnitude spectrum of the output is in theory equal to the magnitude spectrum of the filter. Unfortunately, due to the practical nonwhiteness of the input, perfect or even adequate filter reconstruction proved difficult.

Appendix C

Nonlinear system modeling using Volterra-Wiener functional expansions

Time series analysis generally applies to mathematical and qualitative assessment of a signal recorded from some experimental system. From a systems perspective, such analysis encompasses both system and input description. If the input signal to a system is known however, it is possible to characterize the perhaps nonlinear nature of the system (input/output relationship).

One prominent approach to nonlinear system identification uses the theory of functionals (higher-order functions) developed by Vito Volterra and Maurice Rene Frechet in the late nineteenth century. The theory was extended in Norbert Wiener's seminal work *Cybernetics* to the modeling and characterizations of biological signals.

C.1 Volterra Series Representation

The most common nonlinear system model is known as the Volterra series expansion and is reminiscent of a multi-dimensional Taylor-series expansion:

$$y(n) = k_0 + \sum_{m_1=0}^{\kappa-1} k_1(m_1)x(n-m_1) + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} k_2(m_1,m_2)x(n-m_1)x(n-m_2) + \dots$$

where $x(n)$ is the input signal, $y(n)$ is the output signal, κ the memory of the system, and $k(m_1, m_2, \dots)$ are known as the Volterra kernels, and represent the nature of the system. The Volterra kernels are analogous the impulse response of linear systems theory. A similar equation is used for the modeling of continuous time systems:

$$y(n) = k_0 + \int_0^{\mu} k_1(\tau_1)x(t-\tau_1)d\tau_1 + \int_0^{\mu}\int_0^{\mu} k_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \dots$$

where the kernels are now continuous functions of time and the memory parameter is μ.

The Volterra kernel expansion applies to a specific class of nonlinear systems with the following properties:

1. Time-invariance. As apparent from the expansion, the Volterra kernels are not dependent on a time variable. A straightforward generalization to a time-varying series expansion can be made, but this direction has not been heavily researched.

2. Causality. Only the past and present of the input affect the present of the

output. This restriction can also be generalized away, but as Volterra series analysis is applied principally to temporal signals, causality usually applies.

3. Finite-memory. The Volterra series expansion is best suited for systems with a finite, and preferably short, memory. It is theoretically possible to increase the memory term κ to approximate infinite-memory systems, however, the computational costs for doing so make this approach currently infeasible. The finite memory constraint is analogous to the finite impulse response characteristic from linear systems theory. Consequently, autoregressive (infinite impulse response in the linear case) systems are poorly modeled by the Volterra expansion.

4. Smoothness. Since only the first few Volterra kernels are generally estimated, a practical implication of Volterra kernel estimation assumes some general smoothness criteria for the system to ensure that a polynomial approximation is sound.

C.2 Wiener Series Representation

One major disadvantage to using the Volterra series expansion is that the functionals are not orthogonal [119], namely

$$E[K_i K_j] \neq 0$$

for $i \neq j$, where

$$K_d = \sum_{m_1=0}^{\kappa-1} \sum_{m_2=0}^{\kappa-1} \cdots \sum_{m_d=0}^{\kappa-1} k_d(m_1, m_2, \ldots, m_d) \prod_{\alpha=1}^{d} x(n - m_\alpha)$$

$$y(n) = \sum_{d=0}^{\infty} K_d$$

With uncorrelated input for instance,

$$\begin{aligned} E[K_0 K_2] &= E[k_0 \sum_{m_1=0}^{\kappa-1} \sum_{m_2=0}^{\kappa-1} k_2(m_1, m_2) x(n-m_1) x(n-m_2)] \\ &= k_0 \sum_{m_1=0}^{\kappa-1} \sum_{m_2=0}^{\kappa-1} k_2(m_1, m_2) E[x(n-m_1) x(n-m_2)] \\ &= k_0 \sum_{m_1=0}^{\kappa-1} \sum_{m_2=0}^{\kappa-1} k_2(m_1, m_2) \sigma^2 \delta(m_1 - m_2) \\ &= \sigma^2 k_0 \sum_{m=0}^{\kappa-1} k_2(m, m) \\ &\neq 0 \end{aligned} \tag{C.1}$$

Since the functionals are not orthogonal, estimating one higher-order kernel requires modifying all previously-estimated kernels. Similarly, the truncation error due to estimating a finite number of kernels (finite order) is not minimized. The Wiener series tackles this problem by providing an alternate kernel expansion to the Volterra series with orthogonal functionals (assuming white Gaussian input). The Wiener series is obtained from the Volterra series by Gram-Schmidt orthogonalization [141].

The zeroth and first order functionals of the Wiener series G_d are the same as the zeroth and first order functionals of the Volterra series as they are already orthogonal.

$$G_0 = K_0 = k_0$$

$$G_1 = K_1 = \sum_{m_1=0}^{\kappa-1} k_1(m_1)x(n-m_1)$$

The second order functional is obtained by the first Gram-Schmidt step:

$$\begin{aligned} G_2 &= K_2 - \sigma^2 \sum_{m=0}^{\kappa-1} k_2(m,m) \\ &= \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} k_2(m_1,m_2)x(n-m_1)x(n-m_2) - \sigma^2 \sum_{m=0}^{\kappa-1} k_2(m,m) \end{aligned}$$

Extending this process yields

$$G_r = \sum_{m=0}^{\lfloor r/2 \rfloor} \frac{(-1)^r r!\sigma^{2m}}{(r-2m)!m!2^m} \sum_{m_1=0}^{\kappa-1} \cdots \sum_{m_{r-2m}}^{\kappa-1} \sum_{\lambda_1}^{\kappa-1} \cdots \sum_{\lambda_m}^{\kappa-1}$$

$$k_r(m_1,\ldots,m_{r-2m},\lambda_1,\lambda_1,\ldots,\lambda_m,\lambda_m) \prod_{i=1}^{r-2m} x(n-m_i)$$

The formulae for converting from Volterra to Wiener kernels and vice versa follow:

$$g_r(m_1,\ldots,m_r) = \sum_{m=0}^{\infty} \frac{(r+2m)!\sigma^{2m}}{r!m!2^m} \sum_{\lambda_1}^{\kappa-1} \cdots \sum_{\lambda_m}^{\kappa-1}$$

$$k_{r+2m}(m_1,\ldots,m_r,\lambda_1,\lambda_1,\ldots,\lambda_m,\lambda_m)$$

$$k_r(m_1,\ldots,m_r) = \sum_{m=0}^{\infty} \frac{(-1)^m(r+2m)!\sigma^{2m}}{r!m!2^m} \sum_{\lambda_1}^{\kappa-1} \cdots \sum_{\lambda_m}^{\kappa-1}$$

$$g_{r+2m}(m_1,\ldots,m_r,\lambda_1,\lambda_1,\ldots,\lambda_m,\lambda_m)$$

C.3 Kernel Estimation

Several major approaches exist for Volterra kernel estimation, endorsed by different researchers. One prominent approach is to consider the problem as a ordinary least squares (OLS) regression, namely

$$\vec{y} = \mathbf{X}\vec{k}$$

where $\vec{y} = [y(1), y(2), \ldots, y(N)]^\mathsf{T}$ and $\vec{k} = [k_0, k_1(0), \ldots, k_1(\kappa), k_2(0,0), 2k_2(1,0), k_2(1,1), 2k_2(2,0), 2k_2(2,1), k_2(2,2), \ldots, 2k_2(\kappa, \kappa-1), k_2(\kappa,\kappa)]^\mathsf{T}$.

As expected, OLS applied to such a large system is generally intractable, so several approaches have been proposed to simplify the process. A modification of the Cholesky decomposition known as "fast orthogonalization" was proposed in 1989 [74] followed by a modified approach which supported speedier calculations for higher orders [75]. An autoregressive version of the original method is described in Appendix A.

A maximum likelihood approach to solving the same system of equations was proposed using methods such as gradient descent to find the kernel estimates have also been proposed [38]. Such methods are generally not practical as such methods are often subject to slow convergence rates or local cost minima [95].

An alternate approach estimates the kernels as linear combinations of carefully selected basis functions. A popular choice is the use of orthonormal discrete Laguerre polynomials:

$$b_j(m) = \alpha^{(m-j)/2}\sqrt{1-\alpha}\sum_{k=0}^{j}(-1)^k\binom{m}{k}\binom{j}{k}\alpha^{j-k}(1-\alpha)^k$$

where $b_j(m)$ is the jth order Laguerre polynomial [105, 95, 94]. Individual kernels may now be expressed as coefficients to the linear combination of such polynomials for different combinations (or cross combinations) of j. Details are provided in Reference [95].

Appendix D

Detection of nonlinear dynamics in short, noisy time series

Detection of nonlinear dynamics, a precursor to chaos, has been an elusive problem since the original studies in the field. Methods estimating the dimension and Lyapunov exponent of a time series only prove useful once nonlinear dynamics has been established [131, 108]. A recently published method, based on autoregressive polynomial kernel estimation, solves this problem quite elegantly even for short time series [7, 121].

The Volterra autoregressive method for detection of nonlinear dynamics considers the dynamics of a time series is governed as follows:

$$x(n+1) = k_0 + \sum_{m_1=0}^{\kappa-1} k_1(m_1)x(n-m_1) + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} k_2(m_1,m_2)x(n-m_1)x(n-m_2) + \ldots$$

where $x(n)$ is the time series, κ is the embedding dimension, and $\{k_r(m_1, m_2, \ldots, m_r)\}$

is the set of Volterra kernels, functions characterizing the autoregressive behavior of the dynamical system.

The method creates a series of nonlinear Volterra models with degrees ranging from 1 to $d_{\max}$ and embedding dimensions ranging from 1 to $\kappa_{\max}$, creating $M = (\kappa_{\max} + d_{\max})!/\kappa_{\max}!d_{\max}!$ nonlinear models. This process is then repeated with linear models with degree 1 and embedding dimensions ranging from 1 to M.

For each of the $2M$ models, a prediction error residual is calculated:

$$\epsilon(\kappa, d) = \sqrt{\frac{\sum_{n=1}^{N}(\hat{x}_n(\kappa, d) - x_n)^2}{\sum_{n=1}^{N}(x_n - E[x_n])^2}}$$

and is normalized for overfitting via the Akaike cost function:

$$C(\kappa, d) = \log \epsilon(\kappa, d) + \frac{(\kappa + d)!}{\kappa! d! N}$$

The nonlinear and linear predictors with the lowest cost function are chosen as optimal and then compared. Any statistical test for testing the difference between the means of two distributions may be used and some common choices are the Mann-Whitney rank-sum statistic or the F-test. An example of the application of this method is given in Figure D-1

The autoregressive Volterra method classifies a wide range of time series according to the presence of nonlinear dynamics. Clearly, deterministic chaotic signals are predicted better with a nonlinear model and are detected as nonlinear. The output of systems with finite memory (FIR) driven by noise do nor reject the null hypothesis,

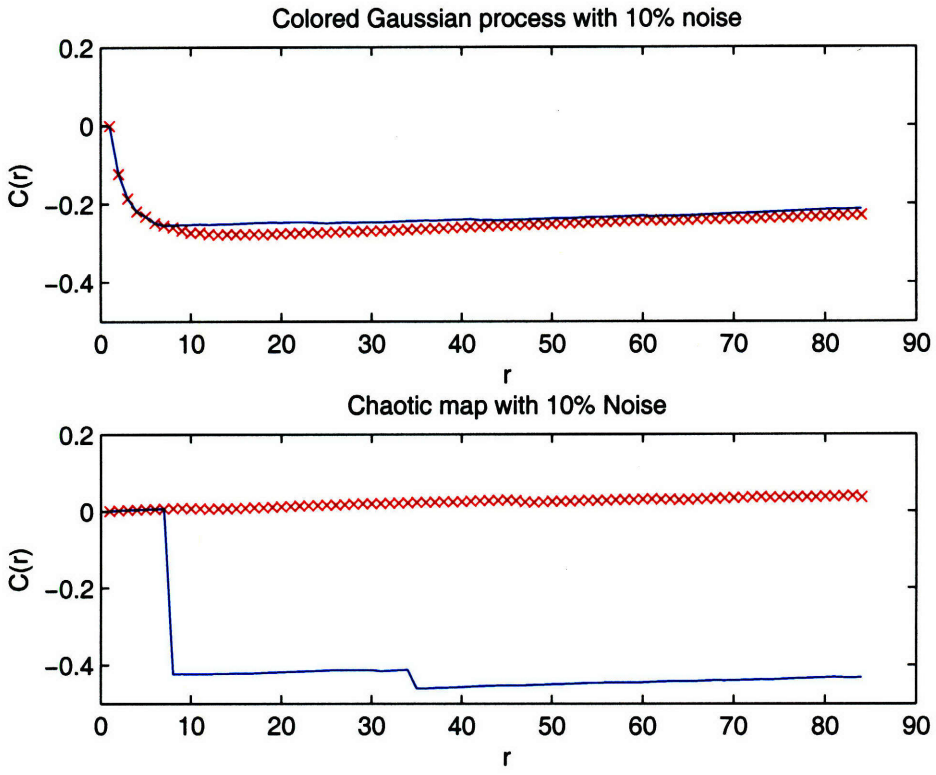

Figure D-1: An example of the autoregressive Volterra method for the detection of nonlinear dynamics in a time series, here applied to noise corrupted colored Gaussian process (top) and a noise corrupted chaotic map (bottom). The nonlinear predictor (blue) is performs significantly better than the linear predictor (red) in the chaotic map and thus nonlinear dynamics are detected. The null hypothesis of linear dynamics are not rejected for the Gaussian process.

as an infinite memory is required for autoregressive dynamical behavior. Uncorrupted periodic signals may either reject or not reject the null hypothesis based on the quality of the respective predictors. Adding noise, however, almost always prefers the linear predictor.

D.1 Orthogonal-search algorithm for efficient autoregressive system estimation

Here we elaborate the method of autoregressive kernel estimation using a modified Cholesky decomposition [74]. The autoregressive Volterra model for a time series is

$$\hat{x}(n+1) = k_0 + \sum_{m_1=0}^{\kappa-1} k_1(m_1)x(n-m_1) + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} k_2(m_1,m_2)x(n-m_1)x(n-m_2) + \ldots$$

where $x(n)$ is the time series, $\hat{x}(n+1)$ is the model prediction, and $k_r(m_1, m_2, \ldots, m_r)$ is the rth Volterra kernel. Our goal is given a delay factor κ and an maximum order d to estimate the kernel functions.

To convert the functional forms of the Volterra kernels into a set of scalar coefficients, we reformulate the expansion as a weighted sum of Volterra terms (or sum of functionals):

$$\hat{x}(n+1) = \sum_{m=0}^{M} a_m p_m(n)$$

where $\{p(n)\}$ is the set of terms in the Volterra series representing all polynomial combinations of the delay terms given a maximum delay κ and maximum order d.

The corresponding set of scalar coefficients are $\{a_m\}$ and represent the kernels. The conversion from the original representation to the modified representation is

$$p_m(n) = \begin{cases} x(n-m+1) & 1 \leq m \leq \kappa \\ x(n-j_1)x(n-j_2) & \kappa+1 \leq m \leq \kappa^2/2 \end{cases}$$

where $j_1 = \lfloor m/\kappa \rfloor$ and $j_2 = mod(m, \kappa)$.

The individual terms $p_m(n)$ are then orthogonalized using the Gram-Schmidt procedure, similar to the orthogonalization process for the Wiener kernel expansion:

$$w_m(n) = p_m(n) - \sum_{r=0}^{m-1} \alpha_{mr} w_r(n)$$

where the coefficients are calculated

$$\alpha_{mr} = \frac{E[p_m(n)w_r(n)]}{E[w_r^2(n)]}$$

The prediction may now be represented as a linear combination of the orthogonalized terms

$$\hat{x}(n+1) = \sum_{m=0}^{M} g_m w_m(n)$$

where

$$g_m = \frac{E[y(n)w_m(n)]}{E[w_m^2(n)]}$$

It follows that the required kernel coefficients can be obtained from g_m:

$$a_m = \sum_{i=m}^{M} g_i v_i$$

where v_i is defined regressively:

$$v_i = \begin{cases} 1 & i = m \\ -\sum_{r=m}^{i-1} \alpha_{ir} v_r & m+1 \leq i \leq M \end{cases}$$

D.2 Implementation of orthogonal search algorithm

To calculate the kernels $\{k_r(m_1, m_2, \ldots, m_r)\}$ for $0 \leq r \leq 2$, we estimate the coefficients in the Gram-Schmidt orthogonalization α_{mr} and the coefficients of the whitened input g_m to obtain scalar kernel terms a_m. Estimating higher-order kernels is analogous.

D.2.1 Estimating α_{mr}

Step 1. Calculate the mean of the time series.

$$\mu_x = E[x(n)] = \frac{1}{N+1} \sum_{n=0}^{N} x(n)$$

Step 2. Calculate the first-, second-, and third-order autocorrelations:

$$\phi_{xx}(i) = E[x(n)x(n-i)] = \frac{1}{N+1} \sum_{n=i}^{N} x(n)x(n-i)$$

$$\phi_{xxx}(i,j) = E[x(n)x(n-i)x(n-j)] = \frac{1}{N+1}\sum_{n=\max(i,j)}^{N} x(n)x(n-i)x(n-j)$$

$$\phi_{xxxx}(i,j,k) = E[x(n)x(n-i)x(n-j)x(n-k)]$$

$$= \frac{1}{N+1)+1}\sum_{n=\max(i,j,k)}^{N} x(n)x(n-i)x(n-j)x(n-k)$$

Step 3. Calculate the means of expansion terms:

$$E[p_m(n)] = \begin{cases} \mu & m = 1 \\ \mu - \frac{1}{N+1}\sum_{i=0}^{m-2} & 2 \le m \le \kappa \\ \phi_{xx}(j_2 - j_1) - \frac{1}{N+1}\sum_{k=0}^{j_1-1} x(k+N-j_1+1)x(k+N-j_2+1) & \kappa+1 \le m \le M \end{cases}$$

where $j_1 = \lfloor m/\kappa \rfloor$ and $j_2 = mod(m,\kappa)$. A recursive scheme may be used to implement these time averages.

Step 4. Calculate the cross-correlations between expansion terms:

$$E[p_m(n)p_r(n)] = \begin{cases} \phi_{xx}(m-1) & r = 1 \\ \phi_{xx}(m-r) - \frac{1}{N+1}\sum_{j=0}^{r-2} x(N-m+2+j)x(N-r+2+j) & \\ & 2 \le r \le m \le \kappa \\ \phi_{xxx}(z_2, z_3) & \\ & z_1 = 0,\ \ 2 \le r \le \kappa < m \le M \\ \phi_{xxx}(z_2 - z_1, z_3 - z_1) & \\ \quad -\frac{1}{N+1}\sum_{k=0}^{z_1-1} x(k+N-z_1+1)x(k+N-z_2+1)x(k+n-z_3+1) & \\ & z_1 \ge 1,\ \ 2 \le r \le \kappa < m \le M \end{cases}$$

where $[z_1, z_2, z_3]$ is the sorted $[j_1, j_2, r-1]$. When $m > \kappa$ and $r > \kappa$, the cross correlation is described algorithmically [74]:

$m = \kappa$

for $j_1 = 0 : \kappa - 1$

  for $j_2 = j_1 : \kappa - 1$

    $m = m + 1$

    for $r = 1 : \kappa$

      $z_1 = \min(j_1, r-1)$

      $A = j_1 + r - 1 - z_1$

      $z_2 = \min(A, j_2)$

      $z_3 = A + j_2 - z_2$

$$q = \sum_{k=0}^{z_1} x(k+N-z_1+1)x(k+N-z_2+1)x(k+N-z_3+1)$$

$$E[p_m(n)p_r(n)] = \phi_{xxx}(z_2 - z_1, z_3 - z_1) - q/(N+1)$$

end

end

end

Step 5. Calculate α_{mr} using the following formula:

$$\alpha_{mr} = \frac{D_{mr}}{B_r}$$

where

$$D_{mr} = \begin{cases} E[p_m(n)] & r = 0 \\ E[p_m(n)p_r(n)] - \sum_{i=0}^{r-1} D_{ri}D_{mi}/B_i & r > 0 \end{cases}$$

$$B_m = \begin{cases} 1 & m = 0 \\ E[p_m^2(n)] - \sum_{r=0}^{m-1} D_{mr}^2/B_r & m > 0 \end{cases}$$

D.2.2 Estimating g_m

Step 1. As with the estimation of α_{mr}, we first calculate the mean and autocorrelations of the time series.

Step 2. Calculate the inner product of the time series with the expansion terms:

$$E[x(n)p_m(n-1)] = \begin{cases} E[x(n)x(n-m)] = \phi_{xx}(m) & 1 \leq m \leq \kappa \\ E[x(n)x(n-1-j_1)x(n-1-j_2)] & \\ \qquad = \phi_{xxx}(j_1+1, j_2+1) & \kappa+1 \leq m \leq M \end{cases}$$

Step 3. Calculate g_m using the following formula:

$$g_m = \frac{C_m}{B_m}$$

where

$$C_m = \begin{cases} \mu_x & m = 0 \\ E[x(n)p_m(n-1)] - \sum_{r=0}^{m-1} \alpha_{mr} C_r & m > 0 \end{cases}$$

D.2.3 Deriving the Volterra kernels

The coefficients of the modified expansion are easily obtainable from α_{mr} and g_m as described previously:

$$a_m = \sum_{i=m}^{M} g_i v_i$$

where

$$v_i = \begin{cases} 1 & i = m \\ -\sum_{r=m}^{i-1} \alpha_{ir} v_r & m+1 \leq i \leq M \end{cases}$$

The kernels of the original representation follow:

$$k_0 = a_0$$

$$k_1(m_1) = a_{m_1+1}$$

$$k_2(m_1, m_2) = k_2(m_2, m_1) = \begin{cases} a_{m_1\kappa+m_2} & m_1 = m_2 \\ a_{m_1\kappa+m_2}/2 & m_1 \neq m_2 \end{cases}$$

Appendix E

Sensitive detection of chaos using noise titration

The detection of chaos given only a time series has been the goal of many studies. Classical methods proved useful for simulated datasets or certain precise physical recordings (described previously), but for a wide class of experimental data especially from physiological and biological sources, noise corruption undermined the efficacy of the classical techniques.

For small amounts of noise, noise reduction methods proved promising, yet still were only applicable if dynamics were presupposed. A novel approach to detection of chaos is to *add* noise to the recorded time series and test for nonlinear dynamics. This approach is known as the method of noise titration for chaos.

Noise titration requires a sensitive test for nonlinear dynamics. The dynamics of nonlinearity are tested on the recorded data series. If linearity is not rejected, then the noise titration method also does not reject the null hypothesis of no chaotic

behavior. If nonlinearity is detected, small amounts of noise (say 1% the standard deviation of the signal) are successively added until nonlinearity is no longer detected. The maximal amount of noise added before nonlinearity is not detected is known as the *noise limit.*

The noise limit acts as both a classification and quantification statistic of the time series. First, if the noise limit is greater than zero this represents the detection of chaotic dynamics. In addition, the noise limit quantifies a measure that ties both the sensitivity to initial conditions (maximal Lyapunov exponent) and the measurement noise corruption of the signal. For noise-free data, the noise limit provides a valuable surrogate for the Lyapunov exponent of a time series [Figure E-1].

E.1 Choice of noise

Noise titration will work with various types of noise, as long as the type of noise corrupts the nonlinearity detection. White Gaussian noise is generally used, but colored or non-Gaussian noise also titrate well. To calculate the efficacy of alternate noise schemes, the average noise limit of the logistic map with $3.5 \leq r \leq 4$ with a step size of 0.002 were tabulated in Table E.1.

E.2 Nonlinear dynamics detection algorithm

It was hypothesized in the original paper [120] that the method of noise titration may be used with any test for nonlinearity. Upon further investigation, it appears the

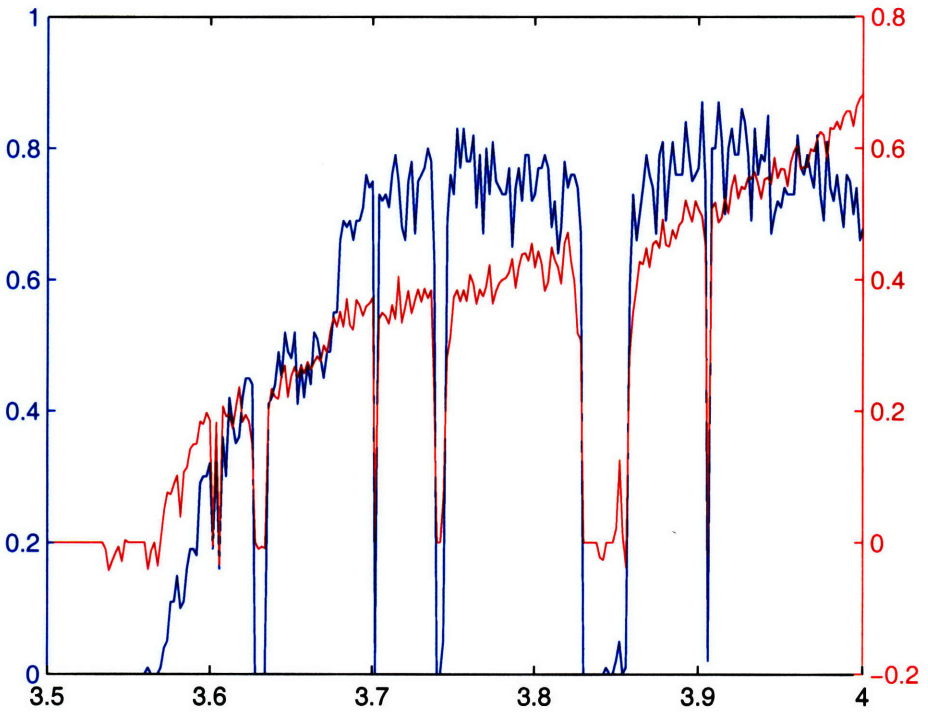

Figure E-1: Noise limit results from titration of chaos (blue) with mathematically calculated Lyapunov exponent values (red) for a range of parameter values for the logistic map.

Signal	Average noise limit
IID Gaussian	0.4978
IID Uniform	0.5249
IID χ^2	0.4829
Differentiated Gaussian	0.3482
Differentiated Uniform	0.2347
Differentiated χ^2	0.4252
LPF Gaussian	0.4722
LPF Uniform	0.3809
LPF χ^2	0.4111

Table E.1: Noise limit comparisons for noise with various distributions and linear filters. The LPF is a exponential moving-average low-pass filter with $\tau = 10$.

autoregressive Volterra algorithm is the only method with sufficient discriminating power for numerical titration. Namely, no other current approach has the ability to cleanly identify nonlinear dynamics. Other methods tried were surrogate data, correlation dimension, Lyapunov estimate, approximate and sample entropies, and a heuristic multifractal measure.

Appendix F

Results from simulations

A wide class of simulated data series were tested to investigate the performance of the three robust algorithms described in the text, the method of surrogate data, the autoregressive Volterra method, and titration of chaos with additive noise.

Several different categories of data were tested, characterized on the type of the system, driven by white Gaussian noise.

- Identity or linear static transform. The output is a white Gaussian process.
- Nonlinear static transform. The output is an IID non-Gaussian process.
- Linear (FIR or IIR) filter. The output is a colored Gaussian process.
- Nonlinear moving-average (finite-memory) filter. A large class of such processes may be broken down into cascades of linear filters and nonlinear static filters:
  - Ln series.
  - nL series.

- Ln parallel.
- LnL series.
- nLn series.
- n series with Ln parallel.
- etc.

In addition, such filters can be created just by finding a nonlinear input-output relation without regards to the linear-nonlinear static decomposition. Fractional moving-average components also belong to this category.

- Nonlinear autoregressive (infinite-memory) filter. The class of deterministic chaos falls into this category. The input Gaussian process is ignored in such a filter.
- Nonlinear autoregressive, moving-average (NARMA) filter. A combination of the above two groups with both autoregressive and moving-average components. The class of systems that exhibit *stochastic chaos* falls into this category.

F.1 Data generation

To test the detection algorithms, we created a large set of each category as follows (R_k is a random real number between -1 and 1, $w(n)$ is the input process, and $x(n)$ is the output process. Each formula was repeated 100 times with different random numbers.):

- Linear static transforms:

$$x(n) = 100R_1 w(n) + 100R_2$$

- Linear filters.

  1. Fractional derivative.

$$X(z) = (1 - z)^{R_1} W(z)$$

     where $X(z)$ and $W(z)$ are the z-transforms of $x(n)$ and $w(n)$ respectively.

  2. Moving-average.

$$x(n) = \sum_{i=1}^{100} R_i x(n + \lfloor 100R_{i+50} \rfloor)$$

  3. Band-pass.

$$X(e^{j\omega}) =< W(e^{j\omega}), [R_1, R_2, \ldots, R_n] >$$

     where $X(e^{j\omega})$ and $W(e^{j\omega})$ are the Fourier transforms of $x(n)$ and $w(n)$ respectively. The $< \cdot, \cdot >$ is the inner product operator.

- Nonlinear static.

  1. Degree-4 polynomial.

$$x(n) = R_1 w(n)^4 + R_2 w(n)^3 + R_3 w(n)^2 + R_4 w(n) + R_5$$

2. Sigmoidal curve.

$$x(n) = \frac{1 + R_1 e^{w(n)/R_2}}{1 + R_3 e^{w(n)/R_4}}$$

3. Random (monotonic) nonlinearity.

$$x(n) = \sum_{i=1}^{w(n)} |R_i|$$

- Nonlinear moving-average.

1. Block structure. For each block structure (Ln, nL, $L|n$, LnL, nLn, $L(L|n)$, $n(L|n)$, $(L|n)L$, $(L|n)n$), choose random linear filters and nonlinear static transforms from above.

2. Logistic MA.

$$x(n) = 4|R_1|w(n + 10R_2)(1 - w(n + 10R_3))$$

3. Henon MA.

$$x(n) = 1 - 2|R_1|w(n + 10R_2)^2 + 2|R_2|w(n + 10R_3)$$

- Nonlinear autoregressive. Such filters have to be carefully designed to prevent divergence. In addition to the random parameter choices, a set of well-known chaotic parameter sets were also used.

1. Logistic map.

$$x(n) = |R_1|x(n-1)(1-x(n-1))$$

2. Cosine map.

$$x(n) = 4|R_1|\cos x(n-1)$$

3. Henon map.

$$x(n) = 1 - 2|R_1|x(n-1)^2 + 2|R_2|x(n-2)$$

4. Ikeda map.

$$x(n) = 1 + R_1(x(n-1)\cos t(n) - y(n-1)\sin t(n))$$

where

$$y(n) = R_1(x(n-1)\sin t(n) + y(n-1)\cos t(n))$$

$$t(n) = \frac{3}{5} - \frac{6}{1 + x(n)^2 + y(n)^2}$$

5. Runge-Kutta discretization of Mackey-Glass flow.

$$\frac{dx(t)}{dt} = \frac{x(t - 100|R_1|)}{5 + 5x(t - 100|R_1|)^{10}} - \frac{x(t)}{10}$$

6. Runge-Kutta discretization of Lorenz flow.

$$\frac{dx(t)}{dt} = 10R_1(y(t) - x(t))$$

$$\frac{dy(t)}{dt} = 10R_2x(t) - y(t) - x(t)z(t)$$

$$\frac{dz(t)}{dt} = x(t)y(t) - R_3z(t)$$

7. Runge-Kutta discretization of Rossler flow.

$$\frac{dx(t)}{dt} = -y(t) - z(t)$$

$$\frac{dy(t)}{dt} = x(t) + R_1y(t)$$

$$\frac{dz(t)}{dt} = R_2 + z(t)(x(t) - 30R_3)$$

- Nonlinear ARMA. Such filters used a random combination of a nonlinear MA and nonlinear AR from above. For iterations with a finite basin of attraction, a wrap-around scheme was used. For example, if a moving-average component pushed the Logistic map to $1 + \epsilon$, the modulus operator would be used to set that value to ϵ and continue. The energy of the MA component relative to the natural energy of the AR component was randomized.

F.2 Results

For each type of data series, 100 realizations of length 2000 were created with different random parameters. For the chaotic maps, commonly investigated parameter choices were also included. These experiments were repeated with 10% measurement noise. The method of surrogate data used 50 surrogates with the z-score test (threshold

3). Approximate entropy and time reversibility, in that order, were used as the discriminating statistics. For the autoregressive Volterra algorithm, parameters of $\kappa = 6$ and $d = 3$ were chosen. Titration rejects the null hypothesis of no chaotic dynamics if the Noise Limit is greater than 2%.

For the flows, integration methods used included the forward Euler, backward Euler, Runge-Kutta, Crank-Nicholson, Richardson extrapolation, and trapezoidal rule. As long as the discretization step size was sufficiently small, we found that the results were a function of the sampling interval and not the integration method of discretization step size. An example of varying sampling interval is shown in Figure F-1. Below, we show the results for the Runge-Kutta integration method with a step size of 0.01. The sampling interval was adjusted until chaos was detected.

- Linear static transforms. Neither the surrogate data method, autoregressive Volterra method, nor titration of chaos rejected the null hypothesis for any of the 100 signals.

- Linear filters. Neither the surrogate data method, autoregressive Volterra method, nor titration of chaos rejected the null hypothesis for any of the 100 signals.

- Nonlinear static. Neither the surrogate data method, autoregressive Volterra method, nor titration of chaos rejected the null hypothesis for any of the 100 signals.

- Nonlinear moving-average. *Ln* processes were rejected by none of the three algorithms. All other block structure constructs and the logistic and Henon MA series were rejected by the method of surrogate data.

- Nonlinear autoregressive. The method of surrogate data rejected the null hypothesis for all time series. The behaviors of the Volterra and titration methods were more complex and is described in Table F.1. The results of the two were similar except for noise-free periodic behavior, which may be detected as nonlinear by Volterra.

- Nonlinear ARMA. The method of surrogate data generally rejected the null hypothesis for all time series. Again the results of the autoregressive Volterra and noise titration methods were a function of the parameter choices and the dynamic noise intensity F.2.

As expected, the dynamic noise component increasingly masked the autoregressive dynamics for the chaotic parameters; as the power of the noise increased, the detection of chaos decreased. Curiously, periodic parameter configurations also now demonstrate chaos with a small amount of dynamic noise. Of course, too much dynamic noise dominates the autoregressive component and chaos is not detected. An example is given in Figure F-2.

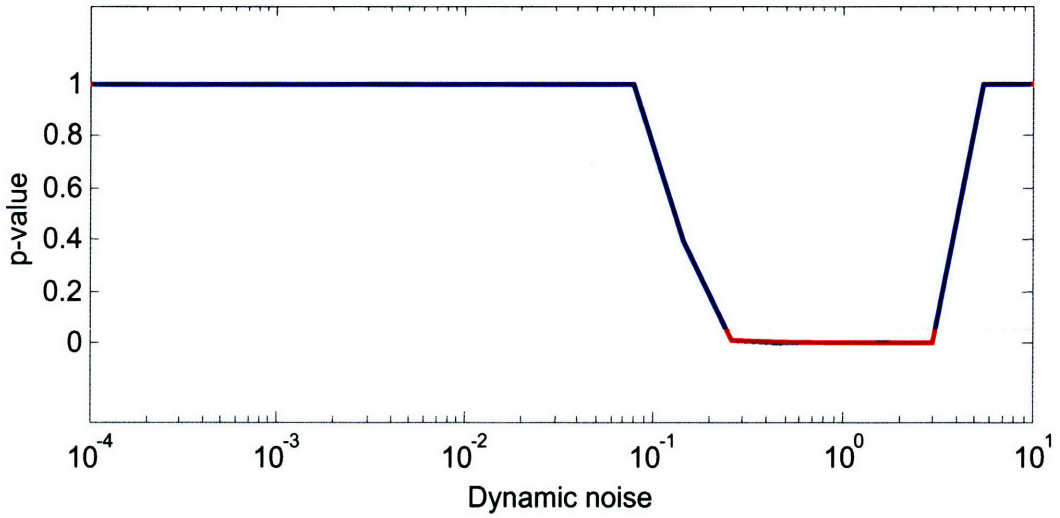

Figure F-1: When discretizing flows, the sampling interval often affects the results. Here we see the VAR p-value calculation results for the Mackey-Glass flow with $\tau = 100$. All series are of length 1000. A p-value less than 0.05 implies detection of nonlinear dynamics. As we see here, a sampling interval too small exaggerates local linear dynamics; whereas, a sampling too large destroys all temporal correlations.

	Volterra	Titration
Periodic	.27	0
Chaotic	1	1
Periodic + Noise	0	0
Chaotic + Noise	.94	.94

Table F.1: The fractions of null hypothesis rejections from the autoregressive Volterra and noise titration methods applied on a randomized collection of nonlinear autoregressive series. Periodicity was tested before the noise was added using an iterative pattern matching scheme on the steady-state components of the time series.

[0% measurement noise]	Volterra	Titration
Periodic parameters + 10% MA	.40	.38
Periodic parameters + 20% MA	.03	0
Periodic parameters + 30% MA	.03	.01
Chaotic parameters + 10% MA	.85	.85
Chaotic parameters + 20% MA	.43	.39
Chaotic parameters + 30% MA	.11	.07
[10% measurement noise]	Volterra	Titration
Periodic parameters + 10% MA	.29	.25
Periodic parameters + 20% MA	0	0
Periodic parameters + 30% MA	0	0
Chaotic parameters + 10% MA	.68	.51
Chaotic parameters + 20% MA	.22	.13
Chaotic parameters + 30% MA	.02	0

Table F.2: The fractions of null hypothesis rejections from the autoregressive Volterra and noise titration methods applied on a randomized collection of nonlinear ARMA series. Periodicity was tested on the MA-free data series. Note that many series autoregressively periodic enter the chaotic regime when driven by dynamical noise.

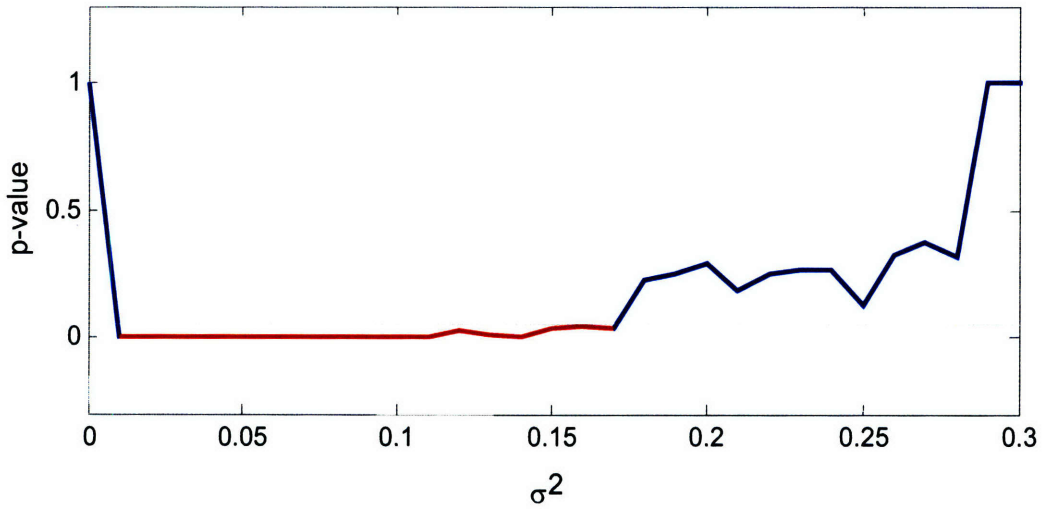

Figure F-2: *LnL* noise added to the logistic map with $r = 3.5$ (with wraparound to prevent divergence). With no dynamic noise, the series is periodic. A small amount of dynamics noise drives the system into chaos; whereas too much masks the nonlinearity. We have found that other moving-average sequences produce similar results.

Appendix G

External Software

External software simplified the testing and development of the methods described in this thesis. Internal software was written using the MATLAB technical computing environment. Certain optimizations and bit-level processing was written in C and interfaced to MATLAB using MEX files.

The following external software packages were used.

- **TISEAN**. An excellent collection of nonlinear time series analysis tools with a focus on nonlinear dynamics by prominent authors of the field [58]. TISEAN runs as binary executables which process ASCII text files of data. In-house software was written to create an interface between TISEAN and MATLAB. [http://www.mpipks-dresden.mpg.de/~tisean/].

- **TSTool**. A lesser known and smaller scale version of TISEAN implemented using MATLAB.
  [http://www.physik3.gwdg.de/tstool/].

- **Michael Small's personal code**. A small collection of MATLAB code, useful for algorithms developed by Small and his coauthors.
  [http://www.eie.polyu.edu.hk/~ensmall/].

- **PhysioNet**. An online collection of recorded physiological signals and software to analyze it [49]. Provides robust implementations of many entropic calculations in addition to code for multifractal analysis using the wavelet formalism.
  [http://www.physionet.org/].

- **LYSIS**. A collection of executable packages for nonlinear system identification by Vasilis Z. Marmarelis, a distinguished researcher in the field. Similar to TISEAN, appropriate adapter software was written for interfacing with MATLAB.
  [http://bmsr.usc.edu/Software/Lysis/lysismenu.html].

Bibliography

[1] LAN Amaral, PC Ivanov, N Aoyagi, I Hidaka, S Tomono, AL Goldberger, HE Stanley, and Y Yamamoto. Behavioral-independent features of complex heartbeat dynamics. *Physical Review Letters*, 86(26):6026–6029, 2001.

[2] A Arneodo, JF Muzy, and S Roux. Experimental Analysis of Self-Similarity and Random Cascade Processes: Application to Fully Developed Turbulence Data. *Journal de Physique II France*, 7:363–370, 1997.

[3] R Badii and A Politi. Statistical description of chaotic attractors: the dimension function. *Journal of Statistical Physics*, 40(5-6):725–750, 1985.

[4] R Badii and A Politi. *Complexity: Hierarchical Structures and Scaling in Physics.* Cambridge University Press, Cambridge, 1999.

[5] JM Balthazar, DT Mook, and JM Rosario. *Nonlinear Dynamics, Chaos, Control, and Their Applications to Engineering Sciences.* Sao Pedro, 1997.

[6] C Bandt and B Pompe. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, 88(17), 2002.

[7] M Barahona and C-S Poon. Detection of nonlinear dynamics in short, noisy time series. *Nature*, 381:16, 1996.

[8] JB Bassingthwaighte. Physiological Heterogeneity: Fractals Link Determinism and Randomness in Structures and Functions. *News in Physiological Sciences*, 3:5–10, 1988.

[9] A Behrman and RJ Baken. Correlation dimension of electroglottographic data from healthy and pathologic subjects. *Journal of the Acoustical Society of America*, 102(4):2371–2379, 1997.

[10] G Boffetta, M Cencini, M Falcioni, and A Vulpiani. Predictability: a way to characterize complexity. *Physics Reports*, 356:367–474, 2002.

[11] S Borovkova, R Burton, and H Dehling. Consistency of the Takens estimator for the correlation dimension. *Annals of Applied Probability*, 9(2):376–390, 1999.

[12] R Bowen and D Ruelle. The ergodic theory of Axiom A flows. *Inventiones Mathematicae*, 29(6):181–202, 1975.

[13] W Brock, WD Dechert, and JA Scheinkman. A test for independence based on the correlation dimension., 1987.

[14] DS Broomhead and GP King. Extracting qualitative dynamics from experimental data. *Physica D*, 20(2-3):217–236, 1986.

[15] R Brown. Calculating Lyapunov exponents for short and/or noisy data sets. *Physical Review E*, 47(6):3962–3969, 1993.

[16] R Brown, P Bryant, and HDI Abarbanel. Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Physical Review A*, 43(6):2787–2806, 1991.

[17] N Burioka, G Cornelissen, F Halberg, and DT Kaplan. Relationship between correlation dimension and indices of linear analysis in both respiratory movement and electroencephalogram. *Clinical Neurophysiology*, 112:1147–1153, 2001.

[18] DC Caccia, D Percival, MJ Cannon, RG Raymond, and JB Bassingthwaigthe. Analyzing exact fractal time seres: evaluating dispersional analysis and rescaled range methods. *Physica A*, 246:609–632, 1997.

[19] MJ Cannon, DB Percival, DC Caccia, GM Raymond, and JB Bassingthwaighte. Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A*, 241(3-4):606–626, 1997.

[20] G Cardano. *The Book on Games of Chance.* Holt, Rinehart and Winston, New York, 1961.

[21] M Casdagli, S Eubank, JD Farmer, and J Gibson. State space reconstruction in the presence of noise. *Physica D*, 51(1):52–98, 1991.

[22] B Castaing, Y Gagne, and EJ Hopfinger. Velocity probability density functions of high Reynolds number turbulence . *Physica D*, 46(2):177–200, 1990.

[23] K-S Chan and H Tong. *Chaos: A Statistical Perspective.* Springer, New York, 2001.

[24] Y Chen and AYT Leung. *Bifurcation and Chaos in Engineering.* Springer, London, 1998.

[25] M Costa, IR Pimentel, T Santiago, P Sarreira, J Melo, and E Ducla-Soares. No evidence of chaos in heart rate variability of normal and cardian transplant human subjects. *Journal of Cardiovascular Electrophysiology*, 10:1350–1357, 1999.

[26] JM Cushing, RF Costantino, B Dennis, RA Desharnais, and SM Henson. *Chaos in Ecology: Experimental Nonlinear Dynamics.* Academic Press, Amsterdam, 2003.

[27] JLR d'Alembert. *Traite de Dynamique.* Gauthier-Villars, Paris, 1921.

[28] M Davies. Noise reduction by gradient descent. *International Journal of Bifurcation and Chaos*, 3:113–118, 1992.

[29] D Delignieres, S Ramdani, L Lemoine, K Torre, M Fortes, and G Ninot. Fractal analyses for 'short' time seres: a re-assessment of classical methods, 2005.

[30] C Diks. Estimating invariants of noisy attractors. *Physical Review E*, 53(5):4263–4266, 1996.

[31] C Diks, JC van Houwelingen, F Takens, and J DeGoede. Reversibility as a criterion for discriminating time series. *Physics Letters A*, 201(2-3):221–228, 1995.

[32] J-P Eckmann and D Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57(3):617–656, 1985.

[33] FY Edgeworth. *Papers relating to political economy.* B Franklin, New York, 1962.

[34] S Ellner. Estimating attractor dimensions from limited data: A new method, with error estimates . *Physics Letters A*, 133(3):128–133, 1988.

[35] IR Epstein and JA Pojman. *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos.* Oxford University Press, New York, 1998.

[36] L Euler. *Elements of Algebra.* Springer-Verlag, New York, 1984.

[37] RA Eve, S Horsfall, and ME Lee. *Chaos, Complexity, and Sociology: Myths, Models, and Theories.* Sage Publications, Thousand Oaks, 1997.

[38] P Eykhoff. *System Idenfication: Paramter and State Estimation.* Wiley, New York, 1974.

[39] JD Farmer and JJ Sidorowich. Predicting chaotic time series. *Physical Review Letters*, 59(8):845–848, 1987.

[40] DP Feldman and JP Crutchfield. Measures of statistical complexity: Why? *Physics Letters A*, 238(4-5):244–252, 1998.

[41] L Flepp, R Holzner, E Brun, M Finardi, and R Badii. Model identification by periodic-orbit analysis for NMR-laser chaos. *Physical Review Letters*, 67(17):2244–2247, 1991.

[42] JBJ Fourier. *The Analytical Theory of Heat.* Dover Publishers, New York, 1955.

[43] M Fraser and HL Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2):1134–1140, 1986.

[44] G Galilei. *Discoveries and opinions of Galileo.* Doubleday, Garden City, 1957.

[45] F Galton. *Natural Inheritance.* Macmillan, London, 1889.

[46] J Gibson, JD Farmer, M Casdagli, and S Eubank. An analytic approach to practical state space reconstruction. *Physica D*, 57(1):1–30, 1991.

[47] L Glass. Chaos and heart rate variability. *Journal of Cardiovascular Electrophysiology*, 10:1358–1360, 1999.

[48] J Gleick. *Chaos: Making a New Science.* Penguin, 1998.

[49] AL Goldberger, LAN Amaral, L Glass, JM Hausdorff, PCh Ivanov, RG Mark, JE Mietus, GB Moody, CK Peng, and HE Stanley. hysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215–e220, 2000.

[50] RB Govindan, K Narayanan, and MS Gopinathan. On the evidence of deterministic chaos in ECG: surrogate and predictability analysis. *Chaos*, 8(2):495–492, 1998.

[51] P Grassberger and I Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9(1-2):189–208, 1983.

[52] HS Greenside, A Wolf, J Swift, and T Pignataro. Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors. *Physical Review A*, 25(6):3453–3456, 1982.

[53] SM Hammel. A noise reduction method for chaotic systems. *Physics Letters A*, 148(8-9):421–428, 1990.

[54] MH Hassoun. *Fundamentals of Artificial Neural Networks.* MIT Press, Cambridge, 1995.

[55] S Haykin. *Neural Networks: A Comprehensive Foundation.* Macmillan, New York, 1994.

[56] RA Heath. *Nonlinear Dynamics: Techniques and Applications in Psychology.* Lawrence Erlbaum Associates, Mahwah, 2000.

[57] R Hegger and H Kantz. Improved false nearest neighbor method to detect determinism in time series data. *Physical Review E*, 60(4):4970–4973, 1999.

[58] R Hegger, H Kantz, and T Schreiber. Practical Implementation of nonlinear time series methods: The TISEAN package. *Chaos*, 9:413–435, 1999.

[59] RC Hilborn. *Chaos and Nonlinear Dynamics.* Oxford, New York, 2000.

[60] M Hirsch and S Smale. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, 1974.

[61] D Hoyer, R Bauer, B Walter, and U Zwiener. Estimation of nonlinear couplings on the basis of complexity and predictability - a new method applied to cardiorespiratory coordination. *IEEE Transactions on Biomedical Engineering*, 45(5):545–562, 1998.

[62] U Hubner, NB Abraham, and CO Weiss. Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH3 laser. *Physical Review A*, 40(11):6354–6365, 1989.

[63] IW Hunter and MJ Korenberg. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 55:135–144, 1986.

[64] C Huygens. *De Ratiociniis in Ludo Aleae.* 157.

[65] K Judd. An improved estimator of dimension and some comments on providing confidence intervals. *Physica D*, 56(2-3):216–228, 1992.

[66] JK Kanters, NH Holstein-Rathlou, and A Agner. Lack of evidence for low-dimensional chaos in heart rate variability. *Journal of Cardiovascular Electrophysiology*, 5(7):591–601, 1994.

[67] H Kantz, J Kurths, and G Mayer-Krees. *Nonlinear Analysis of Physiological Data.* Springer, Berlin, 1998.

[68] H Kantz and T Schreiber. *Nonlinear time series analysis.* Cambridge, Cambridge, 2004.

[69] RA Katz. The Chaos Paradigm: Developments and Applications in Engineering and Science. In *Second Office of Naval Research/NAVPAL Undersea Warfare Center Technical Conference on Nonlinear Dynamics and Full-spectrum Processing*, Mystic, 1994.

[70] RE Kearney and IW Hunter. Nonlinear identification of stretch reflex dynamics. *Annals of Biomedical Engineering*, 16:79–94, 1988.

[71] M Kennel, R Brown, and H Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6):3403–3411, 1991.

[72] J Kepler. *The Harmony of the World.* American Philosophical Society, Philadelphia, 1997.

[73] DE Knuth. *Seminumerical algorithms.* 1998.

[74] MJ Korenberg. Identifying nonlinear difference equation and functional expansion representations: the fast orthogonal algorithm. *Annals of Biomedical Engineering*, 16:123–142, 1988.

[75] MJ Korenberg. Parallel cascade identification and kernel estimation for nonlinear systems. *Annals of Biomedical Engineering*, 19:429–455, 1991.

[76] A Kumar and SK Mullick. Nonlinear dynamical analysis of speech. *Journal of the Acoustical Society of America*, 100(1):615–629, 1996.

[77] M Lakshmanan and S Rajasekar. *Nonlinear Dynamics: Integrability, Chaos, and Patterns.* Springer, New York, 2003.

[78] BP Lathi. *Linear Systems and Signals.* Berkeley-Cambridge Press, Carmichael, 1992.

[79] AL Lavoisier. *Memoir on heat.* N. Watson Academic Publications, New York, 1982.

[80] F Ledrappier and L-S Young. The metric entropy of diffeomorphisms: part II: relations between entropy, exponents, and dimension. *The Annals of Mathematics*, 122(3):540–574, 1985.

[81] YW Lee and M Schetzen. Measurement of the Wiener kernels of a nonlinear system by cross-correlation. *International Journal of Control*, 2(19):237–254, 1965.

[82] K Lehnertz and CE Elger. Can eplileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity. *Physical Review Letters*, 80(22):5019–5022, 1998.

[83] W Liebert and HG Schuster. Proper choice of the time delay for the analysis of chaotic time series. *Physics Letters A*, 142(2-3):107–111, 1989.

[84] LS Liebovitch and T Toth. A fast algorithm to determine fractal dimensions by box counting . *Physics Letters A*, 141(8-9):386–390, 1989.

[85] AW Lo. Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313, 1991.

[86] EN Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[87] V Loreto, G Paladin, and A Vulpiani. Concept of complexity in random dynamical systems. *Physical Review E*, 53(3):2087–2098, 1996.

[88] TH Makikallio, T Seppanen, M Niemela, KEJ Airaksinen, M Tulppo, and HV Huikuri. Abnormalities in beat to beat complexity of heart rate dynamics in patients with a previous myocardial infarction. *Journal of the American College of Cardiology*, 28(4):1005–1011, 1996.

[89] T Malton. *A Compleat Treatise on Perspective, in Theory and Practice.* Robson, Becket, Taylor, Dilly, London, 1779.

[90] B Mandelbrot and J van Ness. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437, 1968.

[91] RN Mantegna and HE Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, Cambridge, 2000.

[92] PZ Marmarelis and VZ Marmarelis. *Advanced Methods of Physiological System Modeling.* Kluwer Academic Publishers, 1994.

[93] PZ Marmarelis and K Naka. White-noise analysis of a neuron chain: an application of the Wiener theory. *Science*, 175(27):1276–1278, 1972.

[94] VZ Marmarelis. Identification of nonlinear biological systems using Laguerre expansions of kernels. *Annals of Biomedical Engineering*, 21:573–589, 1993.

[95] VZ Marmarelis. *Nonlinear Dynamic Modelling of Physiological Systems.* John Wiley and Sons, Hoboken, 2004.

[96] JC Maxwell. Molecules. *Nature*, 8:437–441, 1873.

[97] RM May. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467, 1976.

[98] G Mendel. *Experiments in Plant Hybridisation.* Harvard University Press, Cambridge, 1963.

[99] M Moller, W Lange, F Mitschke, NB Abraham, and U Hubner. Errors from digitizing and noise in estimating attractor dimensions. *Physics Letters A*, 138(4-5):176–182, 1989.

[100] J Munkres. *Topology.* Prentice Hall, 1999.

[101] JF Muzy, E Bacry, and A Arneodo. Wavelets and Multifractal Formalism for Singular Signals: Application to Turbulence Data. *Physical Review Letters*, 67(25):3515–3518, 1991.

[102] R Nagarajan. Surrogate testing of linear feedback processes with non-Gaussian innovations. *Physica A*, 366:530–538, 2006.

[103] I Newton. *The Principia: Mathematical Principles of Natural Philosophy.* University of California Press, 1999.

[104] PL Nunez, BM Wingeier, and RB Silberstein. Spatial-temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks. *Human Brain Mapping*, 13:125–164, 2001.

[105] H Ogura. Estimation of Wiener kernels of a nonlinear system and a fast algorithm using digital Laguerre filters. In *Conference on Information Processing in Neuron Network*, pages 14–62, Okazaki, Japan, 1985.

[106] AV Oppenheim, RW Schafer, and JR Buck. *Discrete Time Signal Processing.* Prentice Hall, New Jersey, 1999.

[107] AV Oppenheim and AS Willsky. *Signals and Systems.* Prentice Hall, 1997.

[108] AR Osborne and A Provenzale. Finite correlation dimension for stochastic systems with power-law spectra. *Physica D*, 35(3):357–381, 1989.

[109] G Paladin, M Serva, and A Vulpiani. Complexity in dynamical systems with noise. *Physical Review Letters*, 74(1):66–69, 1995.

[110] A Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York, 1984.

[111] G Parisi and U Frisch. On the singularity structure of fully developed turbulence, 1985.

[112] U Parlitz. Identification of true and spurious Lyapunov exponents from time series. *International Journal of Bifurcation and Chaos*, 2(1):155–165, 1992.

[113] K Pearson. *Early Statistical Papers.* Cambridge University Press, Cambridge, 1948.

[114] CK Peng, SV Buldyrev, S Havlin, M Simons, and HE Stanley. Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2):1685–1689, 1994.

[115] YB Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys*, 32(4):55–114, 1977.

[116] SM Pincus. Approximate entropy as a measure of system complexity. *PNAS*, 88(6):2297–2301, 1991.

[117] H Poincare. *New Methods in Celestial Mechanics.* Gauthier-Villards, Paris, 1892.

[118] C-S Poon. The chaos about heart rate chaos. *Journal of Cardiovascular Electrophysiology*, 11(2):235–235, 2000.

[119] C-S Poon. Nonlinear systems identification, 2001.

[120] C-S Poon and M Barahona. Titration of chaos with added noise. *PNAS*, 98(13):7107–7112, 2001.

[121] C-S Poon and CK Merrill. Decrease of cardiac chaos in congestive heart failure. *Nature*, 389(6650):492–495, 1997.

[122] WS Pritchard, DW Duke, and KK Krieble. Dimensional analysis of resting human EEG. II: Surrogate-data testing indicates nonlinearity but not low-dimensional chaos. *Psychophysiology*, 32(5):486–491, 1995.

[123] A Provenzale, LA Smith, R Vio, and G Murante. Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D*, 58(1-4):31–49, 1992.

[124] T Puu. *Attractors, Bifurcations, and Chaos: Nonlinear Phenomena in Economics.* Springer, Berling, 2003.

[125] ML Van Quyen, J Martinerie, C Adam, and FJ Varela. Unstable periodic orbits in human epileptic activity. *Physical Review E*, 56(3):3401–3411, 1997.

[126] G Rangarajan and M Ding. Integrated approach to the assessment of long range correlation in time series data. *Physical Review E*, 61(5):4991–5001, 2000.

[127] JS Richman and JR Moorman. Physiological time series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*, 278(6):2039–2049, 2000.

[128] MT Rosenstein, JJ Collins, and CJ De Luca. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D*, 65(1-2):117–134, 1993.

[129] HM Sakai. White-noise analysis in neurophysiology. *Physiological Reviews*, 72(2):491–505, 1992.

[130] T Schreiber. Efficient neighbor searching in nonlinear time series analysis. *International Journal of Bifurcation and Chaos*, 5:349–358, 1995.

[131] T Schreiber and A Schmitz. Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters*, 77(4-22):635–638, 1996.

[132] T Schreiber and A Schmitz. Discrimination power of measures for nonlinearity in a time series. *Physical Review E*, 55(5):5443–5447, 1997.

[133] T Schreiber and A Schmitz. Surrogate time series. *Physica D*, 142(3-4):346–382, 2000.

[134] CE Shannon. *The Mathematical Theory of Communication.* University of Illinoise Press, Urbana, 1964.

[135] M Small, K Judd, M Lowe, and S Stick. Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep. *Journal of Applied Physiology*, 86(1):359–376, 1999.

[136] JC Sprott. *Chaos and time-series analysis.* Oxford, New York, 2004.

[137] RF Sproull. Refinements to nearest-neighbor searching in k-dimensional trees. *Algorithmica*, 9(1):579–589, 1991.

[138] HE Stanley, LAN Amaral, AL Goldberger, S Havlin, PC Ivanov, and CK Peng. Statistical physics and physiology: monofractal and multifractal approaches. *Physica A*, 270:309–324, 1999.

[139] P Stavroulakis. *Chaos Applications in Telecommunications.* CRC, Boca Raton, 2006.

[140] I Stewart. *Does God Play Dice?* Blackwell, Malden, 2002.

[141] G Strang. *Introduction to Linear Algebra.* Wellesley Cambridge Press, 2003.

[142] S Strogatz. *Nonlinear Dynamics and Chaos.* Addison-Wesley, Reading, 1994.

[143] G Sugihara and RM May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268):734–741, 1990.

[144] F Takens. Detecting strange attractors in turbulence. *Physical Review Letters*, 79(8):1475–1478, 1981.

[145] F Takens. On the numerical determination of the dimension of an attractor. In BLJ Braaksma, HW Broer, and F Takens, editors, *Dynamical Systems and Bifurcations, Lecture Notes in Math.*, volume 1125. Springer, Heidelberg-New York, 1985.

[146] J Theiler. Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Physical Review A*, 36(9):4456–4462, 1987.

[147] J Theiler. Some comments on the correlation dimension of $1/f^a$ noise. *Physics Letters A*, 155:480–493, 1991.

[148] J Theiler. Some comments on the correlation dimension of $1/f^\alpha$ noise. *Physics Letters A*, 155:480–493, 1991.

[149] J Theiler, S Eubank, A Longtin, B Galadrikian, and JD Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58(1-4):77–94, 1992.

[150] J Theiler and T Lookman. Statistical error in a chord estimator of correlation dimension: the "rule of five". *International Journal of Bifurcations and Chaos*, 3:765–771, 1993.

[151] G Tononi, O Sporns, and GM Edelman. A measure of brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91:5033–5037, 1994.

[152] DL Turcotte. *Fractals and Chaos in Geology and Geophysics.* Cambridge University Press, Cambridge, 1997.

[153] N Wiener. *Cybernetics; or, Control and Communication in the Animal and the Machine.* MIT Press, New York, 1961.

[154] P Winston. *Artificial Intelligence.* Addison Wesley, 1992.

[155] A Wolf, JB Swift, HL Swinney, and JA Vastano. Determining Lyapunov exponents from a time series. *Physica D*, 16:285–317, 1984.

[156] H Wymeersch. Iterative receiver design, 2006.